



Adaptive Estimation in the Single-Index Model via Oracle Approach

O Lepski, N Serdyukova

► To cite this version:

O Lepski, N Serdyukova. Adaptive Estimation in the Single-Index Model via Oracle Approach. Mathematical Methods of Statistics, 2013, 22 (4), pp.310-322. 10.3103/S1066530713040030 . hal-01265251

HAL Id: hal-01265251

<https://hal.science/hal-01265251>

Submitted on 3 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ADAPTIVE ESTIMATION IN THE SINGLE-INDEX MODEL VIA ORACLE APPROACH *

BY OLEG LEPSKI AND NORA SERDYUKOVA

Aix-Marseille Université and Georg-August-Universität Göttingen

In the framework of nonparametric multivariate function estimation we are interested in structural adaptation. We assume that the function to be estimated has the “single-index” structure where neither the link function nor the index vector is known. We suggest a novel procedure that adapts simultaneously to the unknown index and smoothness of the link function. For the proposed procedure, we prove a “local” oracle inequality (described by the pointwise seminorm), which is then used to obtain the upper bound on the maximal risk of the adaptive estimator under assumption that the link function belongs to a scale of Hölder classes. The lower bound on the minimax risk shows that in the case of estimating at a given point the constructed estimator is optimally rate adaptive over the considered range of classes. For the same procedure we also establish a “global” oracle inequality (under the L_r norm, $r < \infty$) and examine its performance over the Nikol’skii classes. This study shows that the proposed method can be applied to estimating functions of inhomogeneous smoothness, that is whose smoothness may vary from point to point.

1. Introduction. This research aims at estimating multivariate functions with the use of the oracle approach. The first step of the method consists in justification of pointwise and global oracle inequalities for the estimation procedure; the second step is the deriving from them adaptive results for estimation of the point functional and the entire function correspondingly. The obtained results show full adaptivity of the proposed estimator as well as its minimax rate optimality.

Model and set-up. Let $\mathcal{D} \supset [-1/2, 1/2]^d$ be a bounded interval in \mathbb{R}^d . We observe a path $\{Y_\varepsilon(t), t \in \mathcal{D}\}$, satisfying the stochastic differential equation

$$(1.1) \quad Y_\varepsilon(dt) = F(t)dt + \varepsilon W(dt), \quad t = (t_1, \dots, t_d) \in \mathcal{D},$$

where W is a Brownian sheet and $\varepsilon \in (0, 1)$ is the deviation parameter.

*Funding of the ANR-07-BLAN-0234 is acknowledged. The second author is also supported by the DFG FOR 916.

AMS 2000 subject classifications: Primary 62G05; secondary 62G20, 62M99

Keywords and phrases: Adaptive estimation, Gaussian white noise, lower bounds, minimax rate of convergence, nonparametric function estimation, oracle inequalities, single-index, structural adaptation.

In the single-index modeling the signal F has a particular structure:

$$(1.2) \quad F(x) = f(x^\top \theta^\circ),$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is called link function and $\theta^\circ \in \mathbb{S}^{d-1}$ is the index vector.

We consider the case of completely unknown parameters f and θ° and the only technical assumption is that $f \in \mathbb{F}_M$ where $\mathbb{F}_M = \{g : \mathbb{R} \rightarrow \mathbb{R} \mid \sup_{u \in \mathbb{R}} |g(u)| \leq M\}$ for some $M > 0$. However, the knowledge of M as well as any information on the smoothness of the link function are not required for the proposed below estimation procedure. The consideration is restricted to the case $d = 2$ except the second assertion of Theorem 3 concerning a lower bound for function estimation at a given point. Also, without loss of generality we will assume that $\mathcal{D} = [-1, 1]^2$ and $\varepsilon \leq e^{-1}$.

Let $\tilde{F}(\cdot)$ be an estimator, i.e. a measurable function of the observation $\{Y_\varepsilon(t), t \in \mathcal{D}\}$ and \mathbb{E}_F^ε denote the mathematical expectation with respect to \mathbb{P}_F^ε , the family of probability distributions generated by the observation process $\{Y_\varepsilon(t), t \in \mathcal{D}\}$ on the Banach space of continuous functions on \mathcal{D} , when F is the mean function. The estimation quality is measured by the L_r risk, $r \in [1, \infty)$,

$$(1.3) \quad \mathcal{R}_r^{(\varepsilon)}(\tilde{F}, F) = \mathbb{E}_F^\varepsilon \|\tilde{F} - F\|_r,$$

where $\|\cdot\|_r$ is the L_r norm on $[-1/2, 1/2]^2$ or by the “pointwise” risk

$$(1.4) \quad \mathcal{R}_{r,x}^{(\varepsilon)}(\tilde{F}, F) = (\mathbb{E}_F^\varepsilon |\tilde{F}(x) - F(x)|^r)^{1/r}.$$

The aim is to estimate the entire function F on $[-1/2, 1/2]^2$ or its value $F(x)$ from the observation $\{Y_\varepsilon(t), t \in \mathcal{D}\}$ satisfying (1.1) without any prior knowledge of the nuisance parameters f and θ° . More precisely, we will construct an adaptive (not depending of f and θ°) estimator $\hat{F}(x)$ at any point $x \in [-1/2, 1/2]^2$. In what follows \hat{F} notation stands for an adaptive estimator and \tilde{F} denotes an arbitrary estimator. Our estimation procedure is a random selector from a special family of kernel estimators parameterized by a window size (bandwidth) $h > 0$ and a direction of the projection $\theta \in \mathbb{S}^1$, see Section 2.2 below. For this procedure we then establish a pointwise oracle inequality (Theorem 1) of the following type:

$$(1.5) \quad \mathcal{R}_{r,x}^{(\varepsilon)}(\hat{F}, F) \leq C_1 \varepsilon \sqrt{\ln(1/\varepsilon)/h^*(x^\top \theta^\circ)} + C_2 \varepsilon \sqrt{\ln(1/\varepsilon)},$$

where h^* is an optimal in a certain sense (oracle) bandwidths, see Definition 2.1. As $r < \infty$ Jensen’s inequality and Fubini’s theorem trivially imply

$$\left[\mathcal{R}_r^{(\varepsilon)}(\hat{F}, F) \right]^r \leq \mathbb{E}_F^\varepsilon \|\hat{F}(\cdot) - F(\cdot)\|_r^r = \|\mathcal{R}_{r,\cdot}^{(\varepsilon)}(\hat{F}, F)\|_r^r.$$

Hence, we immediately obtain the “global” oracle inequality

$$(1.6) \quad \mathcal{R}_r^{(\varepsilon)}(\hat{F}, F) \leq C_1 \varepsilon \|\sqrt{\ln(1/\varepsilon)/h^*}\|_r + C_2 \varepsilon \sqrt{\ln(1/\varepsilon)}.$$

Both inequalities (1.5) and (1.6) aside of being quite informative itself – we will see in Section 2.1 from Proposition 1 that they claim that our adaptive estimator mimics its ideal (oracle) counterpart, i.e. their risk bounds differ only by a numerical constant, – they are further used to judge the minimax rate of convergence under the pointwise and L_r losses correspondingly (Theorems 3 and 4). We will see that these rates are in accordance with Stone’s dimensionality reduction principle, see pp. 692-693 in Stone (1985). Indeed, as the statistical model is effectively one-dimensional due to the structural assumption (1.2) so the rate of convergence is.

The obtained results demonstrate full adaptivity of the proposed estimator to the unknown direction of the projection θ° and the smoothness of f . Moreover, the lower bound given in the second assertion of Theorem 3 shows that in the case of pointwise estimation over the range of classes of d -variate functions having the single-index structure, see definition (3.1), our estimator is even optimally rate adaptive, that is it achieves the minimax rate of convergence. This fact is in striking contrast to the common knowledge that a payment for pointwise adaptation in terms of convergence rate is unavoidable. Indeed, if the index θ° would be known, than the problem boils down to pointwise adaptation over Hölder classes in the univariate GWN model. As demonstrated in Lepski (1990), an optimally adaptive estimator does not exist in this case.

Although the literature on the single-index model is rather numerous, we mention only books Härdle et al. (2004), Horowitz (1998), Györfi et al. (2002) and Korostelev and Korosteleva (2011), quite a few works address the problem of function estimating when both the link function and index are unknown. To the best of our knowledge the only exceptions are Golubev (1992), Gaïffas and Lecué (2007) and Goldenshluger and Lepski (2008). An adaptive projection estimator is constructed in Golubev (1992), in Gaïffas and Lecué (2007) the aggregation method is used. Both the papers employ L_2 losses. Goldenshluger and Lepski (2008) seems to be the first work on pointwise adaptive estimation in the considered set-up, the upper bound for estimation at a point obtained therein is similar to our, but the estimation procedure is different.

Organization of the paper. In Section 2 we motivate and explain the proposed selection rule. Then in Section 2.3 we establish for it local and global oracle inequalities of type (1.5) and (1.6). In Section 3 we apply these results to minimax adaptive estimation. Particularly, Section 3.1 is devoted to the upper bound and already discussed above lower bound for estimation over a range of Hölder classes. Section 3.2 addresses the “global” adaptation under the L_r losses and the estimator performance over the collection of classes of single-index functions with the link function in a Nikol’skii class, see Definition 2 and (3.2). That consideration allows to incorporate in analysis functions of inhomogeneous smoothness, that is those which can be very smooth on some parts of observation domain and irregular on the others. The proofs of the main results are given in Section 4 and the proofs of technical lemmas are postponed until Appendix.

2. Oracle approach. Below we define an “ideal” (oracle) estimator and describe our estimation procedure. Then we present local and global oracle inequalities demonstrating a nearly oracle performance of the proposed estimator.

Denote by $\mathcal{K} : \mathbb{R} \rightarrow \mathbb{R}$ any function (kernel) that integrates to one, and define for any $z \in \mathbb{R}$, $h \in (0, 1]$ and any $f \in \mathbb{F}_M$

$$\Delta_{\mathcal{K},f}(h, z) = \sup_{\delta \leq h} \left| \frac{1}{\delta} \int \mathcal{K}\left(\frac{u-z}{\delta}\right) [f(u) - f(z)] du \right|,$$

a monotonous approximation error of the kernel smoother $1/\delta \int \mathcal{K}[(u-z)/\delta] f(u) du$. In particular, if the function f is uniformly continuous then $\Delta_{\mathcal{K},f}(h, z) \rightarrow 0$ as $h \rightarrow 0$.

In what follows we assume that the kernel \mathcal{K} obeys

ASSUMPTION 1. (1) $\text{supp}(\mathcal{K}) \subseteq [-1/2, 1/2]$, $\int \mathcal{K} = 1$, \mathcal{K} is symmetric;

(2) there exists $Q > 0$ such that

$$|\mathcal{K}(u) - \mathcal{K}(v)| \leq Q|u - v|, \quad \forall u, v \in \mathbb{R}.$$

2.1. *Oracle estimator.* For any $y \in \mathbb{R}$ denote by

$$\bar{\Delta}_{\mathcal{K},f}(h, y) = \sup_{a>0} \frac{1}{2a} \int_{y-a}^{y+a} \Delta_{\mathcal{K},f}(h, z) dz,$$

the Hardy-Littlewood maximal function of $\Delta_{\mathcal{K},f}(h, \cdot)$, see for instance Wheeden and Zygmund (1977). Put also $\Delta_{\mathcal{K},f}^*(h, \cdot) = \max \{ \bar{\Delta}_{\mathcal{K},f}(h, \cdot), \Delta_{\mathcal{K},f}(h, \cdot) \}$ and remark that in view of the Lebesgue differentiation theorem $\Delta_{\mathcal{K},f}^*(h, \cdot)$ and $\bar{\Delta}_{\mathcal{K},f}(h, \cdot)$ coincide almost everywhere. Note also, that if f is a continuous function then $\Delta_{\mathcal{K},f}^*(h, \cdot) \equiv \bar{\Delta}_{\mathcal{K},f}(h, \cdot)$.

Define for $\forall y \in \mathbb{R}$ the oracle (depending on the underlying function) bandwidth $h_{\mathcal{K},f}^*(y)$

$$(2.1) \quad h_{\mathcal{K},f}^*(y) = \sup \{ h \in [\varepsilon^2, 1] : \sqrt{h} \Delta_{\mathcal{K},f}^*(h, y) \leq \|\mathcal{K}\|_{\infty} \varepsilon \sqrt{\ln(1/\varepsilon)} \}.$$

We see that with the proviso $f \in \mathbb{F}_M$, the “bias” $\Delta_{\mathcal{K},f}^*(h, \cdot) \leq 2M\|\mathcal{K}\|_1$, and consequently the set (2.1) is not empty for all $\varepsilon \leq \exp \{ -(2M\|\mathcal{K}\|_1/\|\mathcal{K}\|_{\infty})^2 \}$. Here $\|\mathcal{K}\|_p$, $1 \leq p \leq \infty$, denotes the L_p norm of \mathcal{K} .

For any $(\theta, h) \in \mathbb{S}^1 \times [\varepsilon^2, 1]$ define the matrix

$$E_{(\theta,h)} = \begin{pmatrix} h^{-1}\theta_1 & h^{-1}\theta_2 \\ -\theta_2 & \theta_1 \end{pmatrix}$$

and consider the family of kernel estimators

$$\mathcal{F} = \left\{ \hat{F}_{(\theta,h)}(\cdot) = \det(E_{(\theta,h)}) \int K(E_{(\theta,h)}(t - \cdot)) Y_{\varepsilon}(dt), (\theta, h) \in \mathbb{S}^1 \times [\varepsilon^2, 1] \right\}.$$

We use the product type kernels $K(u, v) = \mathcal{K}(u)\mathcal{K}(v)$ with a one-dimensional kernel \mathcal{K} obeying Assumption 1. Note also that $\det(E_{(\theta, h)}) = h^{-1}$ and

$$(2.2) \quad \widehat{F}_{(\theta, h)}(\cdot) - \mathbb{E}_F^\varepsilon [\widehat{F}_{(\theta, h)}(\cdot)] \sim \mathcal{N}(0, \|\mathcal{K}\|_2^4 \varepsilon^2 h^{-1}).$$

The choice $\theta = \theta^\circ$ and $h = h^* := h_{\mathcal{K}, f}^*(x^T \theta^\circ)$ leads to the “ideal” (oracle) estimator $\widehat{F}_{(\theta^\circ, h^*)}$, that is the estimator constructed *as if* θ° and f would be known. Such an “estimator” is not available but serves as a quality benchmark, given by the following result.

PROPOSITION 1. *For any $(f, \theta^\circ) \in \mathbb{F}_M \times \mathbb{S}^1$, $\varepsilon \leq \exp\{-\max[1, (2M\|\mathcal{K}\|_1/\|\mathcal{K}\|_\infty)^2]\}$ and any $r \geq 1$*

$$\mathcal{R}_{r, x}^{(\varepsilon)}(\widehat{F}_{(\theta^\circ, h^*)}, F) \leq \mathfrak{c}_r \left[\frac{\|\mathcal{K}\|_\infty^4 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K}, f}^*(x^\top \theta^\circ)} \right]^{1/2}, \forall x \in [-1/2, 1/2]^2,$$

where $\mathfrak{c}_r = [\mathbb{E}(1 + |\varsigma|)^r]^{1/r}$, $\varsigma \sim \mathcal{N}(0, 1)$. The proof is straightforward and can be omitted.

The meaning of Proposition 1 is that the “oracle” knows the exact value of the index θ° and the optimal, up to $\ln(1/\varepsilon)$, bias-variance trade-off h^* between the approximation error caused by $\Delta_{\mathcal{K}, f}^*(h^*, \cdot)$ and the variance, see formula (2.2), of the kernel estimator from the collection \mathcal{F} .

Below we will propose an adaptive (not depending of θ° and f) estimator and show that this estimator is as good as the oracle one, i.e. that the risk of that estimator is worse than that of Proposition 1 by a numerical constant only.

2.2. *Selection rule.* The procedure below is based on a pairwise comparison of the estimators from \mathcal{F} with an auxiliary estimator defined as follows. For any $\theta, \nu \in \mathbb{S}^1$ and any $h \in [\varepsilon^2, 1]$ introduce the matrices

$$\overline{E}_{(\theta, h)(\nu, h)} = \begin{pmatrix} \frac{(\theta_1 + \nu_1)}{2h(1 + |\nu^\top \theta|)} & \frac{(\theta_2 + \nu_2)}{2h(1 + |\nu^\top \theta|)} \\ -\frac{(\theta_2 + \nu_2)}{2(1 + |\nu^\top \theta|)} & \frac{(\theta_1 + \nu_1)}{2(1 + |\nu^\top \theta|)} \end{pmatrix}, \quad E_{(\theta, h)(\nu, h)} = \begin{cases} \overline{E}_{(\theta, h)(\nu, h)}, & \nu^\top \theta \geq 0; \\ \overline{E}_{(-\theta, h)(\nu, h)}, & \nu^\top \theta < 0. \end{cases}$$

It is easy to check that $(4h)^{-1} \leq \det(E_{(\theta, h)(\nu, h)}) \leq (2h)^{-1}$. Then, similarly to the construction of the estimators from \mathcal{F} we define a kernel estimator parametrized by $E_{(\theta, h)(\nu, h)}$

$$(2.3) \quad \widehat{F}_{(\theta, h)(\nu, h)}(x) = \det(E_{(\theta, h)(\nu, h)}) \int K(E_{(\theta, h)(\nu, h)}(t - x)) Y_\varepsilon(dt).$$

Put $\Lambda(\mathcal{K}, Q) = 8\sqrt{\ln(1 + 2Q\|\mathcal{K}\|_\infty)} + 50$ and let for any $\eta \in (0, 1]$

$$\text{TH}(\eta) = 2\|\mathcal{K}\|_\infty^2 [\Lambda(\mathcal{K}, Q) + \sqrt{4r + 2 + 1}] \varepsilon \sqrt{\eta^{-1} \ln(1/\varepsilon)}.$$

Set $\mathcal{H}_\varepsilon = \{h_k = 2^{-k}, k = 0, 1, \dots\} \cap [\varepsilon^2, 1]$ and define for any $\theta \in \mathbb{S}^1$ and $h \in \mathcal{H}_\varepsilon$

$$(2.4) \quad R_{(\theta, h)}(x) = \sup_{\eta \in \mathcal{H}_\varepsilon: \eta \leq h} \left\{ \sup_{\nu \in \mathbb{S}^1} |\hat{F}_{(\theta, \eta)(\nu, \eta)}(x) - \hat{F}_{(\nu, \eta)}(x)| - \text{TH}(\eta) \right\}.$$

For any $x \in [-1/2, 1/2]^2$ introduce the random set

$$\mathcal{P}(x) = \{(\theta, h) \in \mathbb{S}^1 \times \mathcal{H}_\varepsilon : R_{(\theta, h)}(x) \leq 0\},$$

and let $\tilde{h} = \max \{h : (\theta, h) \in \mathcal{P}(x)\}$ if $\mathcal{P}(x) \neq \emptyset$. Note that there exists $\vartheta \in \mathbb{S}^1$ such that $(\vartheta, \tilde{h}) \in \mathcal{P}(x)$, since the set \mathcal{H}_ε is finite. Define

$$\hat{\theta} = \begin{cases} (1, 0)^\top, & \mathcal{P}(x) = \emptyset; \\ \theta \text{ s.t. } (\theta, \tilde{h}) \in \mathcal{P}(x), & \mathcal{P}(x) \neq \emptyset. \end{cases}$$

If $\hat{\theta}$ is not unique, let us make any measurable choice. In particular, if $\hat{\Theta} := \{\theta \in \mathbb{S}^1 : (\theta, \tilde{h}) \in \mathcal{P}(x)\}$ one can choose $\hat{\theta}$ as a vector belonging to $\hat{\Theta}$ with the smallest first coordinate. The measurability of this choice follows from the fact that the mapping $\theta \mapsto R_{(\theta, h)}(x)$ is almost surely continuous on \mathbb{S}^1 . This continuity, in its turn, follows from Assumption 1 (2), bound (5.9) for Dudley's entropy integral proved in Lemma 2 below and the condition $f \in \mathbb{F}_M$. Define

$$(2.5) \quad \hat{h} = \sup \left\{ h \in \mathcal{H}_\varepsilon : \left| \hat{F}_{(\hat{\theta}, h)}(x) - \hat{F}_{(\hat{\theta}, \eta)}(x) \right| \leq \text{TH}(\eta), \forall \eta \leq h, \eta \in \mathcal{H}_\varepsilon \right\}$$

and put as a final estimator $\hat{F}(x) = \hat{F}_{(\hat{\theta}, \hat{h})}(x)$.

The proposed above procedure belongs to the stream of pointwise adaptive procedures originating from Lepski (1990). Indeed, the second step determined by (2.5) for the “frozen” $\hat{\theta}$ is exactly the procedure of Lepski (1990) which was originally developed in the framework of the univariate GWN model. There is a rather vast literature on that topic, we mention Bauer et al. (2009) adapted the method of Lepski (1990) for the choice of the parameter for iterated Tikhonov regularization in nonlinear inverse problems, Bertin and Rivoirard (2009) showed the maxiset optimality of that procedure for bandwidth selection under the *sup* norm losses, Chichignoud (2012) used it for selecting among local bayesian estimators, Gaïffas (2007) studied the problem of pointwise estimation in random design Gaussian regression, Serdyukova (2012) investigated a heteroscedastic Gaussian regression under noise misspecification, among many others.

The application of Lepski (1990) requires some sort of ordering on the set of estimators, for instance in (2.5) as soon as $\hat{\theta}$ is fixed it is due to the monotonicity of the “bias” $\Delta_{\mathcal{K}, f}^*(\cdot, y)$. However, when the projection direction is unknown no natural order on \mathcal{F} is available. This problem is similar to the one arising in generalizations of the pointwise adaptive method for multivariate (anisotropic) settings, see for developments in that direction Lepski and Levit (1999), Kerkycharian et al. (2001) and Goldenshluger and Lepski

(2009). Usually the aforementioned issue requires to introduce an auxiliary estimator and construct a procedure carefully capturing the “incomparability” of the estimators. In the considered set-up it is realized by the first step of procedure with $R_{(\theta,h)}(x)$ given by (2.4).

2.3. Oracle inequalities. Throughout the paper we assume that

$$\varepsilon \leq \exp \left\{ -\max[1, (2M\|\mathcal{K}\|_1/\|\mathcal{K}\|_\infty)^2] \right\}.$$

THEOREM 1. For any $(f, \theta^\circ) \in \mathbb{F}_M \times \mathbb{S}^1$, $x \in [-1/2, 1/2]^2$ and any $r \geq 1$

$$\mathcal{R}_{r,x}^{(\varepsilon)}(\widehat{F}_{(\widehat{\theta}, \widehat{h})}, F) \leq C_{r,1}(Q, \mathcal{K}) \sqrt{\frac{\|\mathcal{K}\|_\infty^4 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(x^T \theta^\circ)}} + C_{r,2}(M, Q, \mathcal{K}) \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}.$$

The constants $C_{r,1}(Q, \mathcal{K})$ and $C_{r,2}(M, Q, \mathcal{K})$ are given in the beginning of the proof.

As already mentioned, the global oracle inequality is obtained by integrating the local oracle inequality. For ease of notation, we write $r(\varepsilon) = C_{r,2}(M, Q, \mathcal{K}) \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}$ and $C_r = C_{r,1}(Q, \mathcal{K})$. It follows from Jensen’s inequality and Fubini’s theorem that

$$\mathcal{R}_r^{(\varepsilon)}(\widehat{F}, F) \leq \|\mathcal{R}_{r,\cdot}^{(\varepsilon)}(\widehat{F}, F)\|_r \leq C_r \left\{ \int_{[-1/2, 1/2]^2} \left[\frac{\|\mathcal{K}\|_\infty^4 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(x^T \theta^\circ)} \right]^{\frac{r}{2}} dx \right\}^{\frac{1}{r}} + r(\varepsilon).$$

Integration by substitution gives:

$$\int_{[-1/2, 1/2]^2} \left[\frac{\|\mathcal{K}\|_\infty^4 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(x^T \theta^\circ)} \right]^{\frac{r}{2}} dx \leq \int_{-1/2}^{1/2} \left[\frac{\|\mathcal{K}\|_\infty^4 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(z)} \right]^{\frac{r}{2}} dz$$

leading to the following result.

THEOREM 2. For any $(f, \theta^\circ) \in \mathbb{F}_M \times \mathbb{S}^1$ and any $r \geq 1$

$$\mathcal{R}_r^{(\varepsilon)}(\widehat{F}_{(\widehat{\theta}, \widehat{h})}, F) \leq C_{r,1}(Q, \mathcal{K}) \left\| \sqrt{\frac{\|\mathcal{K}\|_\infty^4 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(\cdot)}} \right\|_r + C_{r,2}(M, Q, \mathcal{K}) \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}.$$

3. Adaptation. In this section with the use of the local oracle inequality from Theorem 1 we solve the problem of pointwise adaptive estimation over a collection of Hölder classes. Then, we turn to the problem of adaptive estimating the entire function over a collection of Nikol’skii classes with the accuracy of an estimator measured under the L_r risk. That is done with the help of the global oracle inequality given in Theorem 2.

Throughout this section we will assume that the kernel \mathcal{K} satisfies additionally Assumption 2 below. Introduce the following notation: for any $a > 0$ let $m_a \in \mathbb{N}$ be the maximal integer strictly less than a .

ASSUMPTION 2. *There exists $\beta_{\max} > 0$ such that*

$$\int z^j K(z) dz = 0, \quad \forall j = 1, \dots, m_{\beta_{\max}}.$$

3.1. *Pointwise adaptation.* Let us firstly recall the definition of Hölderian functions.

DEFINITION 1. *Let $\beta > 0$ and $L > 0$. A function $g : \mathbb{R} \rightarrow \mathbb{R}$ belongs to the Hölder class $\mathbb{H}(\beta, L)$ if g is m_β -times continuously differentiable, $\|g^{(m)}\|_\infty \leq L$, $\forall m \leq m_\beta$, and*

$$\left| g^{(m_\beta)}(t+h) - g^{(m_\beta)}(t) \right| \leq Lh^{\beta-m_\beta}, \quad \forall t \in \mathbb{R} \text{ and } h > 0.$$

The aim is to estimate the function $F(x)$ at a given point $x \in [-1/2, 1/2]^2$ under the additional assumption that $F \in \mathbb{F}(\beta_{\max}) := \bigcup_{\beta \leq \beta_{\max}} \bigcup_{L > 0} \mathbb{F}_2(\beta, L)$, where

$$(3.1) \quad \mathbb{F}_d(\beta, L) = \left\{ F : \mathbb{R}^d \rightarrow \mathbb{R} \mid F(z) = f(z^\top \theta), \ f \in \mathbb{H}(\beta, L), \ \theta \in \mathbb{S}^{d-1} \right\},$$

$d \geq 2$ is the dimension and β_{\max} is the constant from Assumption 2, which can be arbitrary but must be chosen *a priori*.

THEOREM 3. *Let $\beta_{\max} > 0$ be fixed and let Assumptions 1 and 2 hold. Then, for any $\beta \leq \beta_{\max}$, $L > 0$ and $x \in [-1/2, 1/2]^2$, we have*

$$\sup_{F \in \mathbb{F}_2(\beta, L)} \mathcal{R}_{r,x}^{(\varepsilon)}(\widehat{F}_{(\widehat{\theta}, \widehat{h})}, F) \leq \|\mathcal{K}\|_\infty^2 \left[C_{r,1}(Q, \mathcal{K}) \psi_\varepsilon(\beta, L) + C_{r,2}(L, Q, \mathcal{K}) \varepsilon \sqrt{\ln(1/\varepsilon)} \right],$$

where $\psi_\varepsilon(\beta, L) = L^{\frac{1}{2\beta+1}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{2\beta}{2\beta+1}}$.

Moreover, for any $\beta, L > 0$, $r \geq 1$, $x \in [-1/2, 1/2]^d$ with $d \geq 2$ and any $\varepsilon > 0$ small enough,

$$\inf_{\widetilde{F}} \sup_{F \in \mathbb{F}_d(\beta, L)} \mathcal{R}_{r,x}^{(\varepsilon)}(\widetilde{F}, F) \geq \varkappa \psi_\varepsilon(\beta, L),$$

where infimum is over all possible estimators. Here \varkappa is a numerical constant independent of ε and L .

We conclude that the estimator $\widehat{F}_{(\widehat{\theta}, \widehat{h})}$ is minimax adaptive with respect to the collection of classes $\{\mathbb{F}_d(\beta, L), \ \beta \leq \beta_{\max}, \ L > 0\}$. As already mentioned, this result is quite surprising. Indeed, if for example, the directional vector $\theta = (1, 0)^\top$, i.e. is known, then $\mathbb{F}(\beta, L) = \mathbb{H}(\beta, L)$ and the considered estimation problem can be easily reduced to estimation of f at a given point in the univariate Gaussian white noise model. As it is shown in Lepski (1990) the adaptive estimator over the collection $\{\mathbb{H}(\beta, L), \ \beta \leq \beta_{\max}, \ L > 0\}$ does not exist.

Also, we would like to emphasize that the lower bound result given by the second assertion of the theorem is proved for arbitrary dimension. As to the proof of the first statement of the theorem it is based on the evaluation of the uniform, over $\mathbb{H}_d(\beta, L)$, lower bound for $h_{\mathcal{K},f}^*(\cdot)$ and on the application of Theorem 1. We note also that the upper bound for the minimax risk given in Theorem 3 was earlier given in Goldenshluger and Lepski (2008), but the estimation procedure used there is completely different from our selection rule.

3.2. Adaptive estimation under the L_r losses. We start this section with the definition of the Nikol'skii class of functions.

DEFINITION 2. Let $\beta > 0$, $L > 0$ and $p \in [1, \infty)$ be fixed. A function $g : \mathbb{R} \rightarrow \mathbb{R}$ belongs to the Nikol'skii class $\mathbb{N}_p(\beta, L)$, if g is m_β -times continuously differentiable and

$$\left(\int_{\mathbb{R}} |g^{(m)}(t)|^p dt \right)^{\frac{1}{p}} \leq L, \quad \forall m = 1, \dots, m_\beta;$$

$$\left(\int_{\mathbb{R}} |g^{(m_\beta)}(t+h) - g^{(m_\beta)}(t)|^p dz \right)^{\frac{1}{p}} \leq Lh^{\beta-m_\beta}, \quad \forall h > 0.$$

Later on we assume that $\mathbb{N}_p(\beta, L) = \mathbb{H}(\beta, L)$ if $p = \infty$.

Here the target of estimation is the entire function $F(\cdot)$ under the assumption that $F \in \mathbb{F}_p(\beta_{\max}) := \bigcup_{\beta \leq \beta_{\max}} \bigcup_{L > 0} \mathbb{F}_{2,p}(\beta, L)$, where

$$(3.2) \quad \mathbb{F}_{d,p}(\beta, L) = \left\{ F : \mathbb{R}^d \rightarrow \mathbb{R} \mid F(z) = f(z^\top \theta), f \in \mathbb{N}_p(\beta, L), \theta \in \mathbb{S}^{d-1} \right\}.$$

THEOREM 4. Let $\beta_{\max} > 0$ be fixed and let Assumptions 1 and 2 hold. Then, for any $L > 0$, $p > 1$, $p^{-1} < \beta \leq \beta_{\max}$ and $r \geq 1$,

$$\sup_{F \in \mathbb{F}_{2,p}(\beta, L)} \mathcal{R}_r^{(\varepsilon)}(\hat{F}_{(\hat{\theta}, \hat{h})}, F) \leq \|\mathcal{K}\|_\infty^2 \left[\varkappa C_{r,1}(Q, \mathcal{K}) \varphi_\varepsilon(\beta, L, p) + C_{r,2}(L, Q, \mathcal{K}) \varepsilon \sqrt{\ln(1/\varepsilon)} \right],$$

where

$$\varphi_\varepsilon(\beta, L, p) = \begin{cases} L^{\frac{1}{2\beta+1}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{2\beta}{2\beta+1}}, & (2\beta+1)p > r; \\ L^{\frac{1}{2\beta+1}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{2\beta}{2\beta+1}} [\ln(1/\varepsilon)]^{\frac{1}{r}}, & (2\beta+1)p = r; \\ L^{\frac{1/2-1/r}{\beta-1/p+1/2}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{\beta-1/p+1/r}{\beta-1/p+1/2}}, & (2\beta+1)p < r. \end{cases}$$

The constant \varkappa is independent of ε , L and \mathcal{K} .

Let us make some remarks. First, note that $\mathbb{F}_{2,p}(\beta, L) \supset \mathbb{N}_p(\beta, L)$. Indeed, the class $\mathbb{N}_p(\beta, L)$ can be viewed as the class of functions F satisfying $F(\cdot) = f(\theta^\top \cdot)$ with $\theta = (1, 0)^\top$.

Then, the problem of estimating such (2-variate) functions can be reduced to the estimation of univariate functions observed in the one-dimensional GWN model. In view of this remark the rate of convergence for the latter problem (which can be found for example in Delyon and Juditsky (1996), Donoho et al. (1995)) is the lower bound for the minimax risk defined on $\mathbb{F}_{2,p}(\beta, L)$. Under assumption $\beta p > 1$ this rate of convergence is given by

$$\phi_\varepsilon(\beta, L, p) = \begin{cases} L^{\frac{1}{2\beta+1}} \varepsilon^{\frac{2\beta}{2\beta+1}}, & (2\beta+1)p > r; \\ L^{\frac{1}{2\beta+1}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{2\beta}{2\beta+1}}, & (2\beta+1)p = r; \\ L^{\frac{1/2-1/r}{\beta-1/p+1/2}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{\beta-1/p+1/r}{\beta-1/p+1/2}}, & (2\beta+1)p < r. \end{cases}$$

The minimax rate of convergence in the case $(2\beta+1)p = r$ remains an open problem, and the rate presented in the middle line above is only the lower asymptotic bound for the minimax risk. Therefore the proposed estimator $\hat{F}_{(\hat{\theta}, \hat{h})}$ is adaptive whenever $(2\beta+1)p < r$. In the case $(2\beta+1)p \geq r$ we loose only a logarithmic factor with respect to the optimal rate and, as mentioned in Introduction, the construction of adaptive estimator over a collection $\{\mathbb{F}_{2,p}(\beta, L), \beta > 0, L > 0\}$ in this case remains an open problem.

4. Proofs.

4.1. *Proof of Theorem 1.* The section starts with the constants used in the statement of the theorem as well as technical lemmas whose proofs are postponed to Appendix.

Constants.

$$C_{r,1}(Q, \mathcal{K}) = 8 [\Lambda(\mathcal{K}, Q) + \sqrt{4r+2} + 1] + \mathfrak{c}_r \left[(2 + \sqrt{2})\Lambda(\mathcal{K}, Q) + 2 \right] + 1;$$

$$C_{r,2}(M, Q, \mathcal{K}) = 2^{1/r} [2M + \Lambda(\mathcal{K}, Q)\mathfrak{c}_{2r}].$$

4.1.1. *Auxiliary results.* For any $\theta, \nu \in \mathbb{S}^1$ and $h \in [\varepsilon^2, 1]$ denote

$$\begin{aligned} S_{(\theta, h)(\nu, h)}(x) &= \det(E_{(\theta, h)(\nu, h)}) \int K(E_{(\theta, h)(\nu, h)}(t-x)) F(t) dt, \\ S_{(\theta, h)}(x) &= \det(E_{(\theta, h)}) \int K(E_{(\theta, h)}(t-x)) F(t) dt. \end{aligned}$$

For ease of notation, we write $h_f^* = h_{\mathcal{K}, f}^*(x^\top \theta^\circ)$.

LEMMA 1. *Grant Assumption 1. Then, for any $\nu \in \mathbb{S}^1$ and any $\eta, h \in [\varepsilon^2, 1]$ satisfying*

$\eta \leq h \leq 2^{-1}h_f^*$, one has

$$\begin{aligned} |S_{(\theta^\circ, h)(\nu, h)}(x) - S_{(\nu, h)}(x)| &\leq 2(h_f^*)^{-1/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}; \\ |S_{(\nu, h)}(x) - S_{(\nu, \eta)}(x)| &\leq 2(h_f^*)^{-1/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}; \\ |S_{(\theta^\circ, h)} - F(x)| &\leq (h_f^*)^{-1/2} \|\mathcal{K}\|_\infty \varepsilon \sqrt{\ln(1/\varepsilon)}. \end{aligned}$$

Let $\mathcal{E}_{a,A}$, $0 < a, A < \infty$, be a set of 2×2 matrices such that

$$|\det(E)| \geq a, \quad |E|_\infty \leq A, \quad \forall E \in \mathcal{E}_{a,A}.$$

Here $|E|_\infty = \max_{i,j} |E_{i,j}|$ denotes the supremum norm, the maximum absolute value entry of the matrix E . Later on without loss of generality we will assume that $a \leq A$, $A \geq 1$.

Assume that the function $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}$ is compactly supported on $[-1/2, 1/2]^2$, $\int \mathcal{L} = 1$ and satisfies the Lipschitz condition

$$|\mathcal{L}(u) - \mathcal{L}(v)| \leq \Upsilon |u - v|_2, \quad \forall u, v \in \mathbb{R}^2,$$

where $|\cdot|_2$ is the Euclidian norm. Let $y \in \mathbb{R}^2$ be fixed. On the parameter set $\mathcal{E}_{a,A}$ let a Gaussian random function be defined by

$$\zeta_y(E) = \|\mathcal{L}\|_2^{-1} \sqrt{|\det(E)|} \int \mathcal{L}(E(u - y)) W(du).$$

Put $\mathbf{c}(a, A) = 4\sqrt{2} [\ln(A \vee \{A/a\}^2) + 2 \ln(1 + \sqrt{2}\Upsilon)]^{1/2} + 29$ and $\mathbf{c}_q = (\mathbb{E}(1 + |\varsigma|)^q)^{1/q}$, where $\varsigma \sim \mathcal{N}(0, 1)$.

LEMMA 2. For any $z > 0$

$$\mathbb{P}\left\{ \sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)| \geq \mathbf{c}(a, A) + z \right\} \leq \mathbb{P}\{|\varsigma| \geq z\} \leq e^{-\frac{z^2}{2}}.$$

Moreover, for any $q \geq 1$

$$\left(\mathbb{E} \left[\sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)|^q \right] \right)^{1/q} \leq \mathbf{c}_q \mathbf{c}(a, A)$$

4.1.2. *Proof of Theorem 1.* Let $h^* \in \mathcal{H}_\varepsilon$ be such that $h^* \leq 2^{-1}h_f^* < 2h^*$. Introduce the random events

$$\mathcal{A} = \{(\theta^\circ, h^*) \in \mathcal{P}(x)\}, \quad \mathcal{B} = \{\widehat{h} \geq h^*\}, \quad \mathcal{C} = \mathcal{A} \cap \mathcal{B},$$

and let $\overline{\mathcal{C}}$ denote the event complimentary to \mathcal{C} . We split the proof into two steps.

Risk computation under \mathcal{C} . The triangle inequality gives

$$(4.1) \quad \begin{aligned} \left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - F(x) \right| &\leq \left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - \widehat{F}_{(\widehat{\theta}, h^*)}(x) \right| + \left| \widehat{F}_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\widehat{\theta}, h^*)}(x) \right| \\ &+ \left| \widehat{F}_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\theta^\circ, h^*)}(x) \right| + \left| \widehat{F}_{(\theta^\circ, h^*)}(x) - F(x) \right|. \end{aligned}$$

¹⁰. Since $h^* \geq 4^{-1}h_f^*$ the definition of \widehat{h} yields

$$(4.2) \quad \left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - \widehat{F}_{(\widehat{\theta}, h^*)}(x) \right| 1_{\mathcal{B}} \leq \text{TH}(h^*) \leq \text{TH}(h_f^*/4).$$

Let us make some remarks. Note that $E_{(\theta, h)(\nu, h)} = \pm E_{(\nu, h)(\theta, h)}$ for any θ, ν and h . Hence, we conclude that $\widehat{F}_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(\cdot) \equiv \widehat{F}_{(\widehat{\theta}, h^*)(\theta^\circ, h^*)}(\cdot)$ since \mathcal{K} is symmetric, see Assumption 1. Next, we note that obviously $\mathcal{A} \subseteq \{\mathcal{P}(x) \neq \emptyset\}$ and, moreover, $\mathcal{A} \subseteq \{\widetilde{h} \geq h^*\}$ in view of the definition of \widetilde{h} . Lastly, $(\widehat{\theta}, \widetilde{h}) \in \mathcal{P}(x)$ by definition that means $R_{(\widehat{\theta}, \widetilde{h})}(x) \leq 0$. Consequently,

$$(4.3) \quad \begin{aligned} \left| \widehat{F}_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\theta^\circ, h^*)}(x) \right| 1_{\mathcal{A}} &= \left| \widehat{F}_{(\widehat{\theta}, h^*)(\theta^\circ, h^*)}(x) - \widehat{F}_{(\theta^\circ, h^*)}(x) \right| 1_{\mathcal{A}} \\ &\leq \text{TH}(h^*) \leq \text{TH}(h_f^*/4). \end{aligned}$$

²⁰. Introduce the following notations. For any $\theta, \nu \in \mathbb{S}^1$ and $h \in [\varepsilon^2, 1]$ set

$$\begin{aligned} \xi_{(\theta, h)(\nu, h)}(x) &= \|K\|_2^{-1} \sqrt{\det(E_{(\theta, h)(\nu, h)})} \int K(E_{(\theta, h)(\nu, h)}(t - x)) W(dt); \\ \xi_{(\theta, h)}(x) &= \|K\|_2^{-1} \sqrt{\det(E_{(\theta, h)})} \int K(E_{(\theta, h)}(t - x)) W(dt). \end{aligned}$$

We remark that $|E_{(\theta, h)}|_\infty \leq h^{-1}$ and $|E_{(\theta, h)(\nu, h)}|_\infty \leq h^{-1}$. Moreover,

$$(4.4) \quad (4h)^{-1} \leq \det(E_{(\theta, h)(\nu, h)}) \leq (2h)^{-1}, \quad \det(E_{(\theta, h)}) = h^{-1}.$$

Since $h \in [\varepsilon^2, 1]$, we assert that

$$(4.5) \quad E_{(\theta, h)(\nu, h)}, E_{(\theta, h)} \in \mathcal{E}_{\frac{1}{4}, \frac{1}{\varepsilon^2}}, \quad \forall \theta, \nu \in \mathbb{S}^1, \quad \forall h \in [\varepsilon^2, 1].$$

We note also that for any $\theta, \nu \in \mathbb{S}^1$ and $h \in [\varepsilon^2, 1]$

$$\begin{aligned} \left| \widehat{F}_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\widehat{\theta}, h^*)}(x) \right| &\leq \left| S_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(x) - S_{(\widehat{\theta}, h^*)}(x) \right| \\ &+ \varepsilon \|K\|_2 \sqrt{\det(E_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)})} \left| \xi_{(\theta^\circ, h^*)(\widehat{\theta}, h^*)}(x) \right| + \varepsilon \|K\|_2 \sqrt{\det(E_{(\widehat{\theta}, h^*)})} \left| \xi_{(\widehat{\theta}, h^*)}(x) \right|. \end{aligned}$$

We obtain from the first assertion of Lemma 1 with $\nu = \hat{\theta}$, $h = h^*$, (4.4) and (4.5)

$$\begin{aligned} \left| \hat{F}_{(\theta^\circ, h^*)}(\hat{\theta}, h^*)(x) - \hat{F}_{(\hat{\theta}, h^*)}(x) \right| &\leq \frac{2 \|\mathcal{K}\|_\infty^2}{\sqrt{h_f^*}} \varepsilon \sqrt{\ln(1/\varepsilon)} + \frac{2 + \sqrt{2}}{\sqrt{h_f^*}} \|\mathcal{K}\|_2^2 \varepsilon \sqrt{\ln(1/\varepsilon)} \zeta_\varepsilon(x) \\ (4.6) \qquad \qquad \qquad &\leq \frac{\|\mathcal{K}\|_\infty^2}{\sqrt{h_f^*}} \varepsilon \sqrt{\ln(1/\varepsilon)} [2 + (2 + \sqrt{2}) \zeta_\varepsilon(x)], \end{aligned}$$

where we denoted

$$\zeta_\varepsilon = [\ln(1/\varepsilon)]^{-1/2} \sup_{E \in \mathcal{E}_{\frac{1}{4}, \frac{1}{\varepsilon^2}}} |\zeta_x(E)|.$$

We have also used that $2h^* \leq h_f^* < 4h^*$.

3⁰. We get in view of the third assertion of Lemma 1 that

$$\begin{aligned} \left| \hat{F}_{(\theta^\circ, h^*)}(x) - F(x) \right| &\leq \sqrt{1/h_f^*} \|\mathcal{K}\|_\infty \varepsilon \sqrt{\ln(1/\varepsilon)} + \sqrt{1/h^*} \|\mathcal{K}\|_2^2 \varepsilon |\varsigma| \\ (4.7) \qquad \qquad \qquad &\leq \sqrt{1/h_f^*} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)} (1 + 2|\varsigma|), \end{aligned}$$

where $\varsigma \sim \mathcal{N}(0, 1)$.

4⁰. We obtain from (4.1), (4.2), (4.3), (4.6) and (4.7) and the second assertion of Lemma 2 with $\mathcal{L} = K$, $a = 1/4$, $A = \varepsilon^{-2}$ and $q = r$, noting that $\Upsilon = \sqrt{2}Q\|\mathcal{K}\|_\infty$,

$$\begin{aligned} &\left\{ \mathbb{E} \left| \hat{F}_{(\hat{\theta}, \hat{h})}(x) - F(x) \right|^r 1_{\mathcal{C}} \right\}^{1/r} \\ &\leq 2 \text{TH}(h_f^*/4) + \left[(2 + \sqrt{2}) \Lambda(\mathcal{K}, Q) \mathfrak{c}_r + 2\mathfrak{c}_r + 1 \right] \frac{\|\mathcal{K}\|_\infty^2}{\sqrt{h_f^*}} \varepsilon \sqrt{\ln(1/\varepsilon)} \\ (4.8) \qquad \qquad \qquad &\leq C_{r,1} \frac{\|\mathcal{K}\|_\infty^2}{\sqrt{h_f^*}} \varepsilon \sqrt{\ln(1/\varepsilon)}. \end{aligned}$$

Here we have also used that

$$\sup_{\varepsilon \leq e^{-1}} \left[\frac{4\sqrt{2} \left\{ 2\sqrt{\ln(2/\varepsilon)} + \sqrt{2\ln(1 + 2Q\|\mathcal{K}\|_\infty)} \right\} + 29}{\sqrt{\ln(1/\varepsilon)}} \right] \leq \Lambda(\mathcal{K}, Q).$$

Risk computation under $\bar{\mathcal{C}}$. Since $f \in \mathbb{F}_M$ one can easily evaluate the discrepancy between the adaptive estimator and the value of function

$$\left| \hat{F}_{(\hat{\theta}, \hat{h})}(x) - F(x) \right| \leq M(1 + \|K\|_1) + \varepsilon \|K\|_2 \sqrt{\det(E_{(\hat{\theta}, \hat{h})})} \left| \xi_{(\hat{\theta}, \hat{h})}(x) \right|.$$

We obtain in view of (4.4) and (4.5), taking into account that $\widehat{h} > \varepsilon^2$,

$$\left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - F(x) \right| \leq M(1 + \|\mathcal{K}\|_1^2) + \|\mathcal{K}\|_2^2 \sqrt{\ln(1/\varepsilon)} \zeta_\varepsilon.$$

Thus, applying the second assertion of Lemma 2 with $\mathcal{L} = K$, $a = 1/4$, $A = \varepsilon^{-2}$, $\Upsilon = \sqrt{2}Q\|\mathcal{K}\|_\infty$ and $q = 2r$, we get

$$\left[\mathbb{E}_F^\varepsilon \left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - F(x) \right|^{2r} \right]^{1/2r} \leq [2M + \Lambda(\mathcal{K}, Q)\mathfrak{c}_{2r}] \|\mathcal{K}\|_\infty^2 \sqrt{\ln(1/\varepsilon)}.$$

Here it is used that $1 \leq \|\mathcal{K}\|_1 \leq \|\mathcal{K}\|_2 \leq \|\mathcal{K}\|_\infty$ due to Assumption 1 (I) and that $\varepsilon \leq e^{-1}$.

With $\lambda_r(M, \mathcal{K}, Q) = 2M + \Lambda(\mathcal{K}, Q)\mathfrak{c}_{2r}$ the use of the Cauchy-Schwartz inequality leads to the following bound:

$$(4.9) \quad \left\{ \mathbb{E}_F^\varepsilon \left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - F(x) \right|^r 1_{\overline{\mathcal{C}}} \right\}^{1/r} \lambda_r(M, \mathcal{K}, Q) \|\mathcal{K}\|_\infty^2 \sqrt{\ln(1/\varepsilon)} [\mathbb{P}_F^\varepsilon(\overline{\mathcal{A}}) + \mathbb{P}_F^\varepsilon(\overline{\mathcal{B}})]^{1/2r}.$$

¹⁰. Let us bound from above $\mathbb{P}_F^\varepsilon(\overline{\mathcal{A}})$. We note that

$$(4.10) \quad \begin{aligned} \mathbb{P}_F^\varepsilon(\overline{\mathcal{A}}) &= \mathbb{P}_F^\varepsilon\{(\theta^\circ, h^*) \notin \mathcal{P}(x)\} = \mathbb{P}_F^\varepsilon\{R_{(\theta^\circ, h^*)}(x) > 0\} \\ &\leq \sum_{k: \varepsilon^2 \leq 2^{-k} \leq h^*} \mathbb{P}_F^\varepsilon \left\{ \sup_{\nu \in \mathbb{S}^1} |\widehat{F}_{(\theta^\circ, 2^{-k})(\nu, 2^{-k})}(x) - \widehat{F}_{(\nu, 2^{-k})}(x)| > \text{TH}(2^{-k}) \right\}. \end{aligned}$$

For any k satisfying $2^{-k} \leq h^*$ and any $\nu \in \mathbb{S}^1$, similarly to (4.6), we obtain from the first assertion of Lemma 1 with $h = 2^{-k}$, (4.4) and (4.5) that

$$(4.11) \quad \begin{aligned} \left| \widehat{F}_{(\theta^\circ, 2^{-k})(\nu, 2^{-k})}(x) - \widehat{F}_{(\nu, 2^{-k})}(x) \right| &\leq 2(h_f^*)^{-1/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)} + 2\sqrt{2^k} \|\mathcal{K}\|_2^2 \varepsilon \sqrt{\ln(1/\varepsilon)} \zeta_\varepsilon(x) \\ &\leq 2^{1+k/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)} [1 + \zeta_\varepsilon(x)]. \end{aligned}$$

Here we have also used that $h_f^* \geq 2^{-k}$. Remembering, that

$$\text{TH}(\eta) = 2\|\mathcal{K}\|_\infty^2 [\Lambda(\mathcal{K}, Q) + \sqrt{4r+2} + 1] \varepsilon \sqrt{\eta^{-1} \ln(1/\varepsilon)},$$

we obtain from (4.11) for any k satisfying $2^{-k} \leq h^*$

$$\begin{aligned} &\mathbb{P}_F^\varepsilon \left\{ \sup_{\nu \in \mathbb{S}^1} |\widehat{F}_{(\theta^\circ, 2^{-k})(\nu, 2^{-k})}(x) - \widehat{F}_{(\nu, 2^{-k})}(x)| > \text{TH}(2^{-k}) \right\} \\ &\leq \mathbb{P}_F^\varepsilon \left\{ \sup_{E \in \mathcal{E}_{\frac{1}{4}, \frac{1}{\varepsilon^2}}} |\zeta_x(E)| > \mathfrak{c}(1/4, \varepsilon^{-2}) + \sqrt{(4r+2) \ln(1/\varepsilon)} \right\} \leq \varepsilon^{2r+1}, \end{aligned}$$

in view of the first assertion of Lemma 2. It yields, together with (4.10)

$$(4.12) \quad \mathbb{P}_F^\varepsilon(\overline{\mathcal{A}}) \leq 2\varepsilon^{2r+1} \log_2(1/\varepsilon) \leq 2\varepsilon^{2r}.$$

2^0 . An upper bound on the probability of event $\{\widehat{h} < h^*\}$ is given by

$$(4.13) \quad \begin{aligned} \mathbb{P}_F^\varepsilon(\overline{\mathcal{B}}) &= \mathbb{P}_F^\varepsilon \left(\bigcup_{k: \varepsilon^2 \leq 2^{-k} \leq h^*} \left\{ \left| \widehat{F}_{(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\widehat{\theta}, 2^{-k})}(x) \right| > \text{TH}(2^{-k}) \right\} \right) = \\ &\leq \sum_{k: \varepsilon^2 \leq 2^{-k} \leq h^*} \mathbb{P}_F^\varepsilon \left\{ \left| \widehat{F}_{(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\widehat{\theta}, 2^{-k})}(x) \right| > \text{TH}(2^{-k}) \right\}. \end{aligned}$$

We note that

$$\begin{aligned} \left| \widehat{F}_{(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\widehat{\theta}, 2^{-k})}(x) \right| &\leq \left| S_{(\widehat{\theta}, h^*)}(x) - S_{(\widehat{\theta}, 2^{-k})}(x) \right| \\ &+ \varepsilon \|K\|_2 \sqrt{\det(E_{(\widehat{\theta}, h^*)})} \left| \xi_{(\widehat{\theta}, h^*)}(x) \right| + \varepsilon \|K\|_2 \sqrt{\det(E_{(\widehat{\theta}, 2^{-k})})} \left| \xi_{(\widehat{\theta}, 2^{-k})}(x) \right|. \end{aligned}$$

Applying the second assertion of Lemma 1 with $\nu = \widehat{\theta}$, $h = h^*$, $\eta = 2^{-k}$, (4.4) and (4.5)

$$(4.14) \quad \begin{aligned} \left| \widehat{F}_{(\widehat{\theta}, h^*)}(x) - \widehat{F}_{(\widehat{\theta}, 2^{-k})}(x) \right| &\leq 2(h_f^*)^{-1/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)} + 2\sqrt{2^k} \|\mathcal{K}\|_2^2 \varepsilon \sqrt{\ln(1/\varepsilon)} \zeta_\varepsilon(x) \\ &\leq 2^{1+k/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)} [1 + \zeta_\varepsilon(x)]. \end{aligned}$$

We remark that the right-hand sides of (4.11) and (4.14) coincide and, therefore, repeating the computation led to (4.12) we get

$$(4.15) \quad \mathbb{P}_F^\varepsilon(\overline{\mathcal{B}}) \leq 2\varepsilon^{2r}.$$

We obtain from (4.9), (4.12) and (4.15)

$$(4.16) \quad \left\{ \mathbb{E}_F^\varepsilon \left| \widehat{F}_{(\widehat{\theta}, \widehat{h})}(x) - F(x) \right|^r 1_{\overline{\mathcal{C}}} \right\}^{1/r} \leq 2^{1/r} \lambda_r(M, \mathcal{K}, Q) \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}.$$

The assertion of the theorem follows now from (4.8) and (4.16). ■

4.2. Proof of Theorem 3. We start this section with an auxiliary result used in the proof of the second assertion of the theorem. That result is proved in Kerkycharian et al. (2008), Proposition 7, and for convenience, we formulate it as Lemma 3 below.

4.2.1. *Auxiliary result.* The result cited below concerns a lower bound for estimators of an arbitrary mapping in the framework of GWN model. Below a version adjusted to the estimation at a given point is provided.

Let \mathcal{F} be a nonempty class of functions and let $F : \mathbb{R}^d \rightarrow \mathbb{R}$ be an unknown signal from model (1.1)–(1.2) satisfying $F \in \mathcal{F} \subset \mathbb{L}_2(\mathcal{D})$, $\mathcal{D} = [-1, 1]^d$. The aim is to estimate the functional $F(x)$, $x \in [-1/2, 1/2]^d$.

LEMMA 3. (*Kerkyacharian et al. (2008)*) Assume that for any $\varepsilon > 0$ there exist a positive integer N_ε , $N_\varepsilon \rightarrow \infty$ as $\varepsilon \rightarrow 0$, $\rho \in (0, 1)$, $c > 0$ and functions $F_0, F_1, \dots, F_{N_\varepsilon} \in \mathcal{F}$ such that:

$$(4.17) \quad |F_i(x) - F_0(x)| = \lambda_\varepsilon, \quad \forall i = 1, \dots, N_\varepsilon;$$

$$(4.18) \quad \langle F_i - F_0, F_j - F_0 \rangle \leq c\varepsilon^2 \quad \forall i, j = 1, \dots, N_\varepsilon, i \neq j;$$

$$(4.19) \quad \|F_i - F_0\|_2^2 \leq \rho\varepsilon^2 \ln(N_\varepsilon), \quad \forall i = 1, \dots, N_\varepsilon.$$

Then for $r \geq 1$

$$\inf_{\tilde{F}} \sup_{F \in \mathcal{F}} \left(\mathbb{E}_F^\varepsilon |\tilde{F}(x) - F(x)|^r \right)^{\frac{1}{r}} \geq \frac{1}{2} \left(1 - \sqrt{\frac{e^c - 1}{e^c + 3}} \right) \lambda_\varepsilon.$$

4.2.2. Proof of Theorem 3.

Proof of the first assertion. Under Assumptions 1 and 2 the standard computation of the bias of kernel estimators, for any $f \in \mathbb{H}(\beta, L)$ and any $z \in \mathbb{R}$, gives

$$\Delta_{\mathcal{K},f}(h, z) \leq \frac{Lh^\beta 2^{-\beta} \|\mathcal{K}\|_\infty}{(1 + \beta)m_\beta!} \leq \|\mathcal{K}\|_\infty Lh^\beta.$$

The right-hand side of the latter inequality does not depend of z so

$$\Delta_{\mathcal{K},f}^*(h, z) \leq \|\mathcal{K}\|_\infty Lh^\beta.$$

Hence, $h_{\mathcal{K},f}^*(z) \geq \left(L^{-1} \varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{2/(2\beta+1)}$ for any $z \in \mathbb{R}$ and the first assertion of the theorem follows from Theorem 1.

Proof of the second assertion. The proof is based on the construction of a family $F_0, \dots, F_{N_\varepsilon} \in \mathcal{F} = \mathbb{F}_d(\beta, L) \subset L_2([-1, 1]^d)$ satisfying conditions (4.17)–(4.19) of Lemma 3.

1⁰. Firstly, we construct $F_0, \dots, F_{N_\varepsilon}$ and verify (4.17). Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be such that $\text{supp}(g) \subset (-1/2, 1/2)$, $g \in \mathbb{H}(\beta, 1)$ and $g(0) \neq 0$. Put $h = \left(\mathfrak{a} L^{-1} \varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{2/(2\beta+1)}$, where the constant $\mathfrak{a} > 0$ will be chosen later in order to satisfy (4.19).

For $b > 0$ put $N_\varepsilon = \varepsilon^{-b}$ assuming without loss of generality that N_ε is an integer. The value of b will be determined later in order to satisfy (4.18).

Let $\{\vartheta_i, i = 1, \dots, N_\varepsilon\} \subset \mathbb{S}^{d-1}$ be defined as

$$\vartheta_i = (\theta_i^{(1)}, \theta_i^{(2)}, 0, \dots, 0)^\top, \quad \theta_i^{(1)} = \cos(i/N_\varepsilon), \quad \theta_i^{(2)} = \sin(i/N_\varepsilon),$$

and let $f_i(v) = Lh^\beta g([v - \vartheta_i^\top x]h^{-1})$, $i = 1, \dots, N_\varepsilon$. Finally, set

$$(4.20) \quad F_0 \equiv 0 \quad \text{and} \quad F_i(t) = f_i(\vartheta_i^\top t), \quad i = 1, \dots, N_\varepsilon.$$

Note that F_i , $i = 1, \dots, N_\varepsilon$, obey structural assumption (1.2), with $f = f_i$ and $\theta^\circ = \vartheta_i$. This, together with $g \in \mathbb{H}(\beta, 1)$, allows us to assert that all F_i are in $\mathcal{F} = \mathbb{F}_d(\beta, L)$.

We have, for any $i = 1, \dots, N_\varepsilon$

$$(4.21) \quad \begin{aligned} |F_i(x) - F_0(x)| &= |f_i(\vartheta_i^\top x)| = |g(0)|L^{\frac{1}{2\beta+1}} \left(\mathfrak{a}\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{2\beta}{2\beta+1}} \\ &= |g(0)|\mathfrak{a}^{\frac{2\beta}{2\beta+1}} \psi_\varepsilon(\beta, L), \end{aligned}$$

and, therefore, (4.17) holds with $\lambda_\varepsilon = |g(0)|\mathfrak{a}^{\frac{2\beta}{2\beta+1}} \psi_\varepsilon(\beta, L)$.

2⁰. Now we check (4.18). Set $\theta_{i\perp} = (-\sin(i/N_\varepsilon), \cos(i/N_\varepsilon))$. We have

$$\begin{aligned} \langle F_i, F_j \rangle &= L^2 h^{2\beta} \int_{[-1,1]^d} g(h^{-1}\vartheta_i^\top(t-x))g(h^{-1}\vartheta_j^\top(t-x))dt \\ &\leq 3^{d-2}L^2 h^{2\beta+2} \int_{\mathbb{R}^2} |g(\theta_i^\top u)g(\theta_j^\top u)| du = 3^{d-2}L^2 h^{2\beta+2} |\theta_{i\perp}^\top \theta_j|^{-1} \|g\|_1^2. \\ &= 3^{d-2}L^2 h^{2\beta+2} |\cos(j/N_\varepsilon) \sin(i/N_\varepsilon) - \cos(i/N_\varepsilon) \sin(j/N_\varepsilon)|^{-1} \|g\|_1^2 \\ &= 3^{d-2}L^2 h^{2\beta+2} |\sin((i-j)/N_\varepsilon)|^{-1} \|g\|_1^2 \\ &= 3^{d-2}L^2 h^{2\beta+2} (\sin(|i-j|/N_\varepsilon))^{-1} \|g\|_1^2. \end{aligned}$$

Thus, we obtain

$$(4.22) \quad \begin{aligned} \sup_{i \neq j; i,j=1,\dots,N_\varepsilon} \langle F_i, F_j \rangle &\leq 3^{d-2}L^2 h^{2\beta+2} (\sin(1/N_\varepsilon))^{-1} \|g\|_1^2 \leq 3^{d-2}2L^2 h^{2\beta+2} N_\varepsilon \|g\|_1^2 \\ &= 3^{d-2}2 \|g\|_1^2 \mathfrak{a}^2 \varepsilon^2 \ln(1/\varepsilon) [N_\varepsilon h]. \end{aligned}$$

Hence, choosing $b < 2/(2\beta + 1)$ we conclude that (4.18) holds with any given $c > 0$ for ε small enough.

3⁰. It remains to verify (4.19). By changing variables for any $i = 1, \dots, N_\varepsilon$

$$\|F_i\|_2^2 \leq 3^{d-1} \|g\|_2^2 L^2 h^{2\beta+1} = 3^{d-1} \|g\|_2^2 \mathfrak{a}^2 \varepsilon^2 \ln(1/\varepsilon) = 3^{d-1} \|g\|_2^2 \mathfrak{a}^2 b^{-1} \varepsilon^2 \ln(N_\varepsilon).$$

Here the notation $\|\cdot\|_2$ stands for the L_2 norms on $[-1, 1]^d$ and $[-1/2, 1/2]$ correspondingly. Choosing $\mathfrak{a}^2 = 3^{-db} \|g\|_2^{-2}$ we see that (4.19) is fulfilled with $\rho = 1/3$.

In view of (4.22) the constants c from (4.18) and \mathfrak{a} are chosen independently of L . Thus, the second assertion of the theorem follows from Lemma 3. \blacksquare

4.2.3. Proof of Theorem 4. To prove the theorem we will exploit the ideas developed in Lepski et al. (1997). Moreover, our considerations are, to a great degree, based on the technical result of Lemma 4 below. Its proof is postponed until Appendix.

LEMMA 4. *Grant Assumptions 1 and 2. Then, for any $\mathfrak{p} > 1$, $0 < s \leq \beta_{\max}$, $\mathcal{Q} > 0$,*

$$\sup_{g \in \mathbb{N}_{\mathfrak{p}}(s, \mathcal{Q})} \|\Delta_{\mathcal{K}, g}^*(h, \cdot)\|_{\mathfrak{p}} \leq 2\mathcal{Q}h^s \|\mathcal{K}\|_{\infty} (\tau_{\mathfrak{p}} + 1) [2^{s\mathfrak{p}} - 1]^{-\frac{1}{\mathfrak{p}}}, \quad \forall h > 0.$$

Here $\tau_{\mathfrak{p}}$ is a depending only of \mathfrak{p} constant from the $(\mathfrak{p}, \mathfrak{p})$ -strong maximal inequality.

Proof of Theorem 4. It is suffice to prove the theorem only in the case $r \geq p$. Indeed, remind that the risk $\mathcal{R}_r^{(\varepsilon)}(\cdot, \cdot)$ is described by the L_r norm on $[-1/2, 1/2]$, therefore

$$\mathcal{R}_r^{(\varepsilon)}(\cdot, \cdot) \leq \mathcal{R}_p^{(\varepsilon)}(\cdot, \cdot), \quad r \leq p.$$

Hence the case $r \leq p$ can be reduced to the case $r = p$.

Yet another observation. In view of embedding of Nikol'skii class $\mathbb{N}_p(\beta, L)$ in the Hölder class with parameters $\beta - 1/p$ and cL , $c > 0$, the assumption $\beta p > 1$ provides that $f \in \mathbb{F}_M$ and the assumptions of Theorem 2 are fulfilled. Moreover, in order to obtain the desired the assertion it suffices to bound from above $\left\| \sqrt{\frac{\|\mathcal{K}\|_{\infty}^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K}, f}^*(\cdot)}} \right\|_r$.

Set $\Gamma_0 = \{y \in [-1/2, 1/2] : h_{\mathcal{K}, f}^*(y) = 1\}$ and $\Gamma_k = \{y \in [-1/2, 1/2] : h_{\mathcal{K}, f}^*(y) \in (2^{-k}, 2^{-k+1}] \cap [\varepsilon^2, 1]\}$ for $k = 1, 2, \dots$. Later on, the integration over empty set is supposed to be zero. We have

$$\begin{aligned} \left\| \sqrt{\frac{\|\mathcal{K}\|_{\infty}^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K}, f}^*(\cdot)}} \right\|_r^r &= \int_{\Gamma_0} \left(\sqrt{\frac{\|\mathcal{K}\|_{\infty}^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K}, f}^*(y)}} \right)^r dy \\ &+ \sum_{k \geq 1} \int_{\Gamma_k} \left(\sqrt{\frac{\|\mathcal{K}\|_{\infty}^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K}, f}^*(y)}} \right)^r dy. \end{aligned}$$

The definition of Γ_0 implies

$$(4.23) \quad \int_{\Gamma_0} \left| \sqrt{\frac{\|\mathcal{K}\|_{\infty}^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K}, f}^*(y)}} \right|^r dt \leq [\|\mathcal{K}\|_{\infty}^2 \varepsilon^2 \ln(1/\varepsilon)]^{\frac{r}{2}}.$$

Assumption 1 (2) implies that $\Delta_{\mathcal{K},f}^*(\cdot, y)$ is continuous on $[\varepsilon^2, 1]$, hence for any $k \geq 1$

$$(4.24) \quad \Delta_{\mathcal{K},f}^*(h_{\mathcal{K},f}^*(y), y) = \left[\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(y)} \right]^{\frac{1}{2}}, \quad \forall y \in \Gamma_k.$$

Let $0 \leq q_k \leq r$ be a sequence whose choice will be done later. We obtain from (4.24)

$$(4.25) \quad \begin{aligned} & \sum_{k \geq 1} \int_{\Gamma_k} \left(\sqrt{\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{h_{\mathcal{K},f}^*(y)}} \right)^r dy \\ & \leq \sum_{k \geq 1} \left(\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{2^{-k}} \right)^{\frac{r-q_k}{2}} \int_{\Gamma_k} \left(\Delta_{\mathcal{K},f}^*(2^{1-k}, y) \right)^{q_k} dy \\ & \leq \sum_{k \geq 1} \left(\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{2^{-k}} \right)^{\frac{r-q_k}{2}} \int \left(\Delta_{\mathcal{K},f}^*(2^{1-k}, y) \right)^{q_k} dy =: \Xi. \end{aligned}$$

To get the first inequality we have used that $\Delta_{\mathcal{K},f}^*(\cdot, y)$ is monotonically increasing function.

The computation of the quantity on the right-hand side of (4.25), including the choice of $(q_k, k \geq 1)$, will be done differently in dependence on β, p and r . Later on c_1, c_2, \dots , denote constants independent on ε, L and \mathcal{K} .

1⁰. *Case* $(2\beta + 1)p > r$. Put

$$h^* = \left[L^{-2} \varepsilon^2 \ln(1/\varepsilon) \right]^{\frac{1}{2\beta+1}}$$

and choose $q_k = p$ if $2^{-k} \leq h^*$ and $q_k = 0$ if $2^{-k} > h^*$.

Applying Lemma 4 with $\mathbf{p} = p$, $s = \beta$ and $\mathcal{Q} = L$ we get

$$(4.26) \quad \begin{aligned} \Xi & \leq c_1 (L \|\mathcal{K}\|_\infty)^p \sum_{k: 2^{-k} \leq h^*} \left(\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{2^{-k}} \right)^{\frac{r-p}{2}} 2^{-k\beta p} + c_2 \left(\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{h^*} \right)^{\frac{r}{2}} \\ & \leq c_3 \|\mathcal{K}\|_\infty^r \left[L^p (\varepsilon^2 \ln(1/\varepsilon))^{\frac{r-p}{2}} \sum_{k: 2^{-k} \leq h^*} 2^{-k[\beta p - \frac{r-p}{2}]} + \left(\frac{\varepsilon^2 \ln(1/\varepsilon)}{h^*} \right)^{\frac{r}{2}} \right]. \end{aligned}$$

Because in the considered case $\beta p - \frac{r-p}{2} > 0$, we obtain

$$\Xi \leq c_4 \|\mathcal{K}\|_\infty^r \left[L^p (\varepsilon^2 \ln(1/\varepsilon))^{\frac{r-p}{2}} (h^*)^{\beta p - \frac{r-p}{2}} + \left(\frac{\varepsilon^2 \ln(1/\varepsilon)}{h^*} \right)^{\frac{r}{2}} \right].$$

It remains to note that h^* is chosen by balancing two terms on the right-hand side of the latter inequality. It yields

$$(4.27) \quad \Xi \leq 2c_4 \left[\|\mathcal{K}\|_\infty L^{\frac{1}{2\beta+1}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{2\beta}{2\beta+1}} \right]^r.$$

The argument in the case $(2\beta + 1)p > r$ is completed with the use of Theorem 2, (4.23) and (4.27).

2⁰. *Case* $(2\beta + 1)p = r$. Put $h^* = 1$ and choose $q_k = p$ for all $k \geq 1$. Repeating the computations led to (4.26) we get

$$(4.28) \quad \Xi \leq c_5 \ln(1/\varepsilon) \left[\|\mathcal{K}\|_\infty L^{p/r} (\varepsilon^2 \ln(1/\varepsilon))^{\frac{r-p}{2r}} \right]^r.$$

Here we have used that $\beta p - \frac{r-p}{2} = 0$ and that the summation in (4.25) runs over k such that $2^{-k} \geq \varepsilon^2$, since otherwise $\Gamma_k = \emptyset$. It remains to note that the equality $(2\beta + 1)p = r$ is equivalent to $p/r = 1/(2\beta + 1)$ and $(r - p)/2r = \beta/(2\beta + 1)$. The assertion of the theorem in the case $(2\beta + 1)p = r$ follows now from Theorem 2, (4.23) and (4.28).

3⁰. *Case* $(2\beta + 1)p < r$. Choose $q_k = r$ if $2^{-k} \leq h^*$ and $q_k = p$ if $2^{-k} > h^*$, where the choice of h^* will be done later.

The following embedding holds, see Besov et al. (1979): $\mathbb{N}_p(\beta, L) \subseteq \mathbb{N}_r(\beta - 1/p + 1/r, c_6 L)$. Thus, applying Lemma 4 with $\mathbf{p} = r$, $s = \beta - 1/p + 1/r$ and $\mathcal{Q} = c_6 L$ we get

$$(4.29) \quad \begin{aligned} \Xi_1 &:= \sum_{k: 2^{-k} \leq h^*} \left(\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{2^{-k}} \right)^{\frac{r-q_k}{2}} \int \left(\overline{\Delta}_{\mathcal{K},f}(2^{1-k}, y) \right)^{q_k} dy \\ &= \sum_{k: 2^{-k} \leq h^*} \int \left(\overline{\Delta}_{\mathcal{K},f}(2^{1-k}, y) \right)^r dy \leq c_7 (\|\mathcal{K}\|_\infty L)^r (h^*)^{\beta r - (r/p) + 1}. \end{aligned}$$

Applying Lemma 4 with $\mathbf{p} = r$, $s = \beta$ and $\mathcal{Q} = L$ we get

$$(4.30) \quad \begin{aligned} \Xi_2 &:= \sum_{k: 2^{-k} > h^*} \left(\frac{\|\mathcal{K}\|_\infty^2 \varepsilon^2 \ln(1/\varepsilon)}{2^{-k}} \right)^{\frac{r-q_k}{2}} \int \left(\overline{\Delta}_{\mathcal{K},f}(2^{1-k}, y) \right)^{q_k} dy \\ &= c_8 L^p (\|\mathcal{K}\|_\infty)^r (\varepsilon^2 \ln(1/\varepsilon))^{\frac{r-p}{2}} \sum_{k: 2^{-k} > h^*} 2^{-k} \left[\beta p - \frac{r-p}{2} \right] \\ &\leq c_9 L^p (\|\mathcal{K}\|_\infty)^r (\varepsilon^2 \ln(1/\varepsilon))^{\frac{r-p}{2}} (h^*)^{\beta p - \frac{r-p}{2}}. \end{aligned}$$

Here we have used that $\beta p - \frac{r-p}{2} < 0$. In view of (4.29) and (4.30) we choose h^* from the equality:

$$L^r (h^*)^{\beta r - (r/p) + 1} = L^p (\varepsilon^2 \ln(1/\varepsilon))^{\frac{r-p}{2}} (h^*)^{\beta p - \frac{r-p}{2}}.$$

It yields $h^* = \left(L^{-1} \varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{1}{\beta - 1/p + 1/2}}$ and we obtain finally that

$$(4.31) \quad \Xi \leq c_{10} (\|\mathcal{K}\|_\infty)^r L^{\frac{r(1/2 - 1/r)}{\beta - 1/p + 1/2}} \left(\varepsilon \sqrt{\ln(1/\varepsilon)} \right)^{\frac{r(\beta - 1/p + 1/r)}{\beta - 1/p + 1/2}}.$$

The assertion of the theorem in the case $(2\beta + 1)p < r$ follows now from Theorem 2, (4.23) and (4.31). \blacksquare

5. Appendix.

5.1. Proof of Lemma 1.

Proof of the first assertion. The symmetry of the kernel \mathcal{K} (Assumption 1 (1)) implies

$$S_{(-\theta^\circ, h)(\nu, h)}(\cdot) \equiv S_{(\theta^\circ, h)(-\nu, h)}(\cdot), \quad S_{(-\nu, h)}(\cdot) \equiv S_{(\nu, h)}(\cdot).$$

Therefore it suffices to prove the first assertion of the lemma under the condition $\nu^\top \theta^\circ \geq 0$. In this case $E_{(\theta^\circ, h)(\nu, h)} = \bar{E}_{(\theta^\circ, h)(\nu, h)}$ and we note that

$$(5.1) \quad \bar{E}_{(\theta^\circ, h)(\nu, h)} = \left[E_{(\theta^\circ, h)}^{-1} + E_{(\nu, h)}^{-1} \right]^{-1}.$$

For any $\theta = (\theta_1, \theta_2) \in \mathbb{S}^1$ let $\theta_\perp = (-\theta_2, \theta_1)$. Using (5.1) we obtain

$$\begin{aligned} S_{(\theta^\circ, h)(\nu, h)}(x) &= \int K(u) f(h[\theta^\circ + \nu]^\top \theta^\circ u_1 + [\theta^\circ_\perp + \nu_\perp]^\top \theta^\circ u_2 + x^\top \theta^\circ) du \\ &= \int \int \mathcal{K}(u_1) \mathcal{K}(u_2) f(h[1 + \nu^\top \theta^\circ] u_1 + \nu_\perp^\top \theta^\circ u_2 + x^\top \theta^\circ) du_1 du_2. \end{aligned}$$

We also have

$$S_{(\nu, h)}(x) = \int \int \mathcal{K}(u_1) \mathcal{K}(u_2) f(h\nu^\top \theta^\circ u_1 + \nu_\perp^\top \theta^\circ u_2 + x^\top \theta^\circ) du_1 du_2.$$

Put $S_\nu^*(x) = \int \mathcal{K}(u_2) f(\nu_\perp^\top \theta^\circ u_2 + x^\top \theta^\circ) du_2$ and consider two cases.

1⁰. $\nu_\perp^\top \theta^\circ = 0$. In this case $S_\nu^*(x) = f(x^\top \theta^\circ)$ and

$$\begin{aligned} S_{(\nu, h)}(x) &= \int \mathcal{K}(u_1) f(hu_1 + x^\top \theta^\circ) du_1 = h^{-1} \int \mathcal{K}([t - x^\top \theta^\circ]/h) f(t) dt, \\ S_{(\theta^\circ, h)(\nu, h)}(x) &= \int \mathcal{K}(u_1) f(2hu_1 + x^\top \theta^\circ) du_1 = (2h)^{-1} \int \mathcal{K}([t - x^\top \theta^\circ]/2h) f(t) dt. \end{aligned}$$

Here we have used that $\nu_\perp^\top \theta^\circ = 0$ together with $\nu^\top \theta^\circ \geq 0$ implies $\nu = \theta^\circ$. Thus, we obtain

$$\begin{aligned} |S_{(\theta^\circ, h)(\nu, h)}(x) - S_{(\nu, h)}(x)| &\leq |S_{(\theta^\circ, h)(\nu, h)}(x) - S_\nu^*(x)| + |S_{(\nu, h)}(x) - S_\nu^*(x)| \\ (5.2) \quad &\leq \Delta_{\mathcal{K}, f}(h, x^\top \theta^\circ) + \Delta_{\mathcal{K}, f}(2h, x^\top \theta^\circ) \leq 2\Delta_{\mathcal{K}, f}^*(2h, x^\top \theta^\circ). \end{aligned}$$

2⁰. $\nu^\top \theta^\circ \neq 0$. In this case we have

$$S_\nu^*(x) = \int \int \frac{1}{h(1 + \nu^\top \theta^\circ)} \mathcal{K} \left(\frac{v_1}{h(1 + \nu^\top \theta^\circ)} \right) \frac{1}{|\nu_\perp^\top \theta^\circ|} \mathcal{K} \left(\frac{v_2 - x^\top \theta^\circ}{|\nu_\perp^\top \theta^\circ|} \right) f(v_2) dv_1 dv_2,$$

$$S_{(\theta^\circ, h)(\nu, h)}(x) = \int \int \frac{1}{h(1 + \nu^\top \theta^\circ)} \mathcal{K} \left(\frac{v_1}{h(1 + \nu^\top \theta^\circ)} \right) \frac{1}{|\nu_\perp^\top \theta^\circ|} \mathcal{K} \left(\frac{v_2 - x^\top \theta^\circ}{|\nu_\perp^\top \theta^\circ|} \right) f(v_1 + v_2) dv_1 dv_2.$$

Here we have used once again the symmetry of kernel \mathcal{K} . Thus, taking into account that $|\nu^\top \theta^\circ| \leq 1$, we get

$$\begin{aligned} & |S_{(\theta^\circ, h)(\nu, h)}(x) - S_\nu^*(x)| \\ & \leq \int \frac{1}{|\nu_\perp^\top \theta^\circ|} \left| \mathcal{K} \left(\frac{v_2 - x^\top \theta^\circ}{|\nu_\perp^\top \theta^\circ|} \right) \right| \sup_{\delta \leq 2h} \left| \int \frac{1}{\delta} \mathcal{K} \left(\frac{v_1}{\delta} \right) [f(v_1 + v_2) - f(v_2)] dv_1 \right| dv_2 \\ & \leq \|\mathcal{K}\|_\infty \sup_{a > 0} \left[\frac{1}{a} \int_{x^\top \theta^\circ - a/2}^{x^\top \theta^\circ + a/2} \sup_{\delta \leq 2h} \left| \int \frac{1}{\delta} \mathcal{K} \left(\frac{v_1}{\delta} \right) [f(v_1 + v_2) - f(v_2)] dv_1 \right| dv_2 \right]. \end{aligned}$$

Here we have used that $\text{supp}(\mathcal{K}) \subseteq [-1/2, 1/2]$ (Assumption 1 (1)). Hence,

$$(5.3) \quad |S_{(\theta^\circ, h)(\nu, h)}(x) - S_\nu^*(x)| \leq \|\mathcal{K}\|_\infty \bar{\Delta}_{\mathcal{K}, f}(2h, x^\top \theta^\circ) \leq \|\mathcal{K}\|_\infty \Delta_{\mathcal{K}, f}^*(2h, x^\top \theta^\circ).$$

If $\nu^\top \theta^\circ \neq 0$ we obtain by the same computations

$$|S_{(\nu, h)}(x) - S_\nu^*(x)| \leq \|\mathcal{K}\|_\infty \Delta_{\mathcal{K}, f}^*(h, x^\top \theta^\circ).$$

Noting that $S_{(\nu, h)}(\cdot) \equiv S_\nu^*(\cdot)$ if $\nu^\top \theta^\circ = 0$ we get

$$(5.4) \quad |S_{(\nu, h)}(x) - S_\nu^*(x)| \leq \|\mathcal{K}\|_\infty \Delta_{\mathcal{K}, f}^*(h, x^\top \theta^\circ),$$

that yields together with (5.3)

$$(5.5) \quad |S_{(\theta^\circ, h)(\nu, h)}(x) - S_{(\nu, h)}(x)| \leq 2\|\mathcal{K}\|_\infty \Delta_{\mathcal{K}, f}^*(2h, x^\top \theta^\circ).$$

Finally, taking into account that in view of Assumption 1 (1) $\|\mathcal{K}\|_\infty \geq 1$, we obtain from (5.2) and (5.5) that

$$|S_{(\theta^\circ, h)(\nu, h)}(x) - S_{(\nu, h)}(x)| \leq 2\|\mathcal{K}\|_\infty \Delta_{\mathcal{K}, f}^*(2h, x^\top \theta^\circ) \leq 2\|\mathcal{K}\|_\infty \Delta_{\mathcal{K}, f}^*(h_f^*, x^\top \theta^\circ),$$

since we consider h such that $2h \leq h_f^*$. The definition of h_f^* implies

$$\Delta_{\mathcal{K}, f}^*(h_f^*, x^\top \theta^\circ) \leq (h_f^*)^{-1/2} \|\mathcal{K}\|_\infty \varepsilon \sqrt{\ln(1/\varepsilon)}$$

and the first assertion of the lemma follows.

Proof of the second and third assertions. In view of (5.4) for $\forall \eta \leq h \leq h_f^*$

$$\begin{aligned} |S_{(\nu,\eta)}(x) - S_{(\nu,h)}(x)| &\leq |S_{(\nu,\eta)}(x) - S_\nu^*(x)| + |S_{(\nu,h)}(x) - S_\nu^*(x)| \\ &\leq \|\mathcal{K}\|_\infty \left[\Delta_{\mathcal{K},f}^*(\eta, x^\top \theta^\circ) + \Delta_{\mathcal{K},f}^*(h, x^\top \theta^\circ) \right] \leq 2\|\mathcal{K}\|_\infty \Delta_{\mathcal{K},f}^*(h, x^\top \theta^\circ) \\ &\leq 2\|\mathcal{K}\|_\infty \Delta_{\mathcal{K},f}^*(h_f^*, x^\top \theta^\circ) \leq 2(h_f^*)^{-1/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}, \end{aligned}$$

in view of the definition of h_f^* . The second assertion is proved.

We have for any $h \leq h_f^*$

$$\begin{aligned} |S_{(\theta^\circ, h)}(x) - F(x)| &= \left| \frac{1}{h} \int \mathcal{K}\left(\frac{u}{h}\right) [f(u + x^\top \theta^\circ) - f(x^\top \theta^\circ)] du \right| \\ &\leq \Delta_{\mathcal{K},f}(h, x^\top \theta^\circ) \leq \Delta_{\mathcal{K},f}^*(h, x^\top \theta^\circ) \leq \Delta_{\mathcal{K},f}^*(h_f^*, x^\top \theta^\circ) \\ &= (h_f^*)^{-1/2} \|\mathcal{K}\|_\infty^2 \varepsilon \sqrt{\ln(1/\varepsilon)}, \end{aligned}$$

in view of the definition of h_f^* . The third assertion is proved. ■

5.2. *Proof of Lemma 2.* Since $\zeta_y(\cdot)$ is a zero mean Gaussian random function we have

$$(5.6) \quad \mathbb{P} \left\{ \sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)| \geq u \right\} \leq 2\mathbb{P} \left\{ \sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E) \geq u \right\}, \quad \forall u > 0.$$

By Lemma 12.2 in Lifshits (1995) the median m of the random variable $\sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E)$ is dominated by the expectation, that is $m \leq \mathbb{E} \sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E)$. That along with the Borell, Tsirelson, Sudakov concentration inequality, see Theorem 12.2 in Lifshits (1995) provides

$$(5.7) \quad \mathbb{P} \left\{ \sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E) \geq \mathbb{E} \sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E) + z \right\} \leq \mathbb{P} \left\{ \sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E) \geq m + z \right\} \leq \mathbb{P} \{ \varsigma \geq z \}$$

since $\sup_{E \in \mathcal{E}_{a,A}} \text{Var}[\zeta_y(E)] = 1$. Here $\varsigma \sim \mathcal{N}(0, 1)$. Thus, to complete the proof of the first assertion of the lemma it suffices to bound $\mathbb{E} \sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E)$. This will be done by the application of Dudley's theorem, see Theorem 14.1 in Lifshits (1995). Denote by ϱ the semi-metric generated by $\zeta_y(\cdot)$ on $\mathcal{E}_{a,A}$:

$$\varrho(E, E') = \sqrt{\mathbb{E} |\zeta_y(E) - \zeta_y(E')|^2}, \quad E, E' \in \mathcal{E}_{a,A}.$$

Without loss of generality one can assume that $|\det(E)| \geq |\det(E')|$, then we have

$$\begin{aligned}
\varrho^2(E, E') &= 2 \left[1 - \|\mathcal{L}\|_2^{-2} \sqrt{|\det(E)| |\det(E')|} \int \mathcal{L}(Ev) \mathcal{L}(E'v) dv \right] \\
&= 2 \left[1 - \|\mathcal{L}\|_2^{-2} \sqrt{\frac{|\det(E')|}{|\det(E)|}} \int \mathcal{L}(z) \mathcal{L}(E'E^{-1}z) dz \right] \\
&= 2 \left[1 - \|\mathcal{L}\|_2^{-2} \sqrt{\frac{|\det(E')|}{|\det(E)|}} \int_{[-\frac{1}{2}, \frac{1}{2}]^2} \mathcal{L}(z) \mathcal{L}(E'E^{-1}z) dz \right] \\
&= 2 \left[1 - \frac{\sqrt{|\det(E')|}}{\sqrt{|\det(E)|}} \right] + \frac{2}{\|\mathcal{L}\|_2^2} \frac{\sqrt{|\det(E')|}}{\sqrt{|\det(E)|}} \left[\int_{[-\frac{1}{2}, \frac{1}{2}]^2} \mathcal{L}(z) [\mathcal{L}(z) - \mathcal{L}(E'E^{-1}z)] dz \right].
\end{aligned}$$

One bounds the first summand with the use of the assumption $|\det(E)| \geq a$:

$$2 \left[1 - \frac{\sqrt{|\det(E')|}}{\sqrt{|\det(E)|}} \right] \leq \frac{2}{\sqrt{a}} \left| \sqrt{|\det(E)|} - \sqrt{|\det(E')|} \right| \leq \frac{2}{\sqrt{a}} |\det(E) - \det(E')|^{1/2}.$$

As for the second term, putting

$$\mathfrak{d}^2(E, E') = \int_{[-\frac{1}{2}, \frac{1}{2}]^2} |\mathcal{L}(E'E^{-1}z) - \mathcal{L}(z)|^2 dz,$$

by the Cauchy-Schwartz inequality we get

$$\int_{[-\frac{1}{2}, \frac{1}{2}]^2} \mathcal{L}(z) [\mathcal{L}(z) - \mathcal{L}(E'E^{-1}z)] dz \leq \|\mathcal{L}\|_2 \mathfrak{d}(E, E').$$

As $\|\mathcal{L}\|_2 \geq 1$, we have

$$\varrho^2(E, E') \leq 2a^{-1/2} |\det(E) - \det(E')|^{1/2} + 2\mathfrak{d}(E, E').$$

First, we note that

$$|\det(E) - \det(E')| \leq 4A|E - E'|_\infty.$$

Second, because \mathcal{L} satisfies the Lipschitz condition with a constant Υ , we have

$$\mathfrak{d}(E, E') \leq \Upsilon \sup_{z \in [-\frac{1}{2}, \frac{1}{2}]^2} |(E' - E)E^{-1}z|_2 \leq 2\sqrt{2}\Upsilon Aa^{-1}|E - E'|_\infty.$$

Since we assumed $a \leq A$, the following bound holds:

$$(5.8) \quad \varrho^2(E, E') \leq 4(\sqrt{2}\Upsilon + 1)Aa^{-1} \left(|E - E'|_\infty^{1/2} \vee |E - E'|_\infty \right).$$

Consider the cube $[0, A]^4$ endowed with the vector supremum norm $|z|_\infty = \max_{i=1,\dots,4} |z_i|$. Let $\mathfrak{E}_{[0,A]^4, |\cdot|_\infty}(\cdot)$ denote the metric entropy of $[0, A]^4$ measured in $|\cdot|_\infty$. Then

$$\mathfrak{E}_{[0,A]^4, |\cdot|_\infty}(\epsilon) \leq 4 \ln(A) + [4 \ln(1/(2\epsilon))]_+, \quad \forall \epsilon \in (0, 1].$$

Denoting by $\mathfrak{E}_{\mathcal{E}_{a,A}, \varrho}(\cdot)$ the metric entropy of $\mathcal{E}_{a,A}$ measured in ϱ , we get in view of (5.8)

$$\mathfrak{E}_{\mathcal{E}_{a,A}, \varrho}(\delta) \leq \mathfrak{E}_{[0,A]^4, |\cdot|_\infty} \left(\frac{\delta^4}{16(1 + \sqrt{2}\Upsilon)^2 A^2 a^{-2}} \right), \quad \forall \delta \in (0, 1],$$

and, therefore,

$$\mathfrak{E}_{\mathcal{E}_{a,A}, \varrho}(\delta) \leq 4 \left[\ln(A \vee \{A/a\}^2) + \ln 8 + 2 \ln(1 + \sqrt{2}\Upsilon) + 4 \ln(1/\delta) \right].$$

Since $\sup_{E \in \mathcal{E}_{a,A}} \text{Var}[\zeta_y(E)] = 1$ the use of Dudley's integral bound, see Theorem 14.1 in Lifshits (1995), leads to

$$\begin{aligned} \mathbb{E} \left[\sup_{E \in \mathcal{E}_{a,A}} \zeta_y(E) \right] &\leq 4\sqrt{2} \int_0^{1/2} \sqrt{\mathfrak{E}_{\mathcal{E}_{a,A}, \varrho}(\delta)} \, d\delta \\ (5.9) \quad &\leq 4\sqrt{2} \left[\ln(A \vee \{A/a\}^2) + 2 \ln(1 + \sqrt{2}\Upsilon) \right]^{1/2} + 29 =: \mathbf{c}(a, A). \end{aligned}$$

Here we have used that $\int_0^{1/2} \sqrt{\ln(1/\delta)} \, d\delta \leq 2^{-1} \sqrt{\pi}$. The first assertion of the lemma follows now from (5.6), (5.7), (5.9) and the standard bound for the Gaussian tail.

To justify the second assertion we first note that for any $q \geq 1$

$$\mathbb{E} \left[\sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)| \right]^q = q \int_0^\infty u^{q-1} \mathbb{P} \left(\sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)| \geq u \right) du.$$

Hence, applying the first assertion of the lemma we have

$$\begin{aligned} \mathbb{E} \left[\sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)| \right]^q &\leq [\mathbf{c}(a, A)]^q + q \int_0^\infty \mathbb{P} \{ |\varsigma| \geq z \} (\mathbf{c}(a, A) + z)^{q-1} dz \\ &= \mathbb{E} (\mathbf{c}(a, A) + |\varsigma|)^q, \end{aligned}$$

where $\varsigma \sim \mathcal{N}(0, 1)$. Thus, we finally have

$$\left(\mathbb{E} \left[\sup_{E \in \mathcal{E}_{a,A}} |\zeta_y(E)| \right]^q \right)^{1/q} \leq \mathbf{c}_q \mathbf{c}(a, A).$$

■

5.3. *Proof of Lemma 4.* First, in view of the (\mathbf{p}, \mathbf{p}) -strong maximal inequality, see e.g. Theorem 9.16 in Wheeden and Zygmund (1977), one has

$$\|\overline{\Delta}_{\mathcal{K},g}(h, \cdot)\|_{\mathbf{p}} \leq \tau_{\mathbf{p}} \|\Delta_{\mathcal{K},g}(h, \cdot)\|_{\mathbf{p}},$$

where the constant $\tau_{\mathbf{p}}$ depends only of \mathbf{p} . Since $\Delta_{\mathcal{K},g}^*(h, \cdot) \leq \overline{\Delta}_{\mathcal{K},g}(h, \cdot) + \Delta_{\mathcal{K},g}(h, \cdot)$ we have

$$(5.10) \quad \|\Delta_{\mathcal{K},g}^*(h, \cdot)\|_{\mathbf{p}} \leq (\tau_{\mathbf{p}} + 1) \|\Delta_{\mathcal{K},g}(h, \cdot)\|_{\mathbf{p}}.$$

For any $\delta \in (0, h]$ put $B(z, \delta) = \left| \delta^{-1} \int \mathcal{K}([u - z]/\delta) (g(u) - g(z)) du \right|$ and define

$$\Delta_{\mathcal{K},g}^{(n)}(h, z) = \sup_{\delta \in [hn^{-1}, h]} B(z, \delta), \quad n = 1, 2, \dots$$

We remark that the sequence $\{\Delta_{\mathcal{K},g}^{(n)}(h, \cdot)\}_{n \geq 1}$ increases monotonically and $\Delta_{\mathcal{K},g}^{(n)}(h, z) \rightarrow \Delta_{\mathcal{K},g}(h, z)$ for any $z \in \mathbb{R}$, as $n \rightarrow \infty$. Hence, by Beppo-Levi's theorem

$$\|\Delta_{\mathcal{K},g}(h, \cdot)\|_{\mathbf{p}} = \lim_{n \rightarrow \infty} \left\| \Delta_{\mathcal{K},g}^{(n)}(h, \cdot) \right\|_{\mathbf{p}},$$

and, in view of (5.10), to complete the argument we need to show that

$$(5.11) \quad \sup_{g \in \mathbb{N}_{\mathbf{p}}(s, \mathcal{Q})} \left\| \Delta_{\mathcal{K},g}^{(n)}(h, \cdot) \right\|_{\mathbf{p}} \leq 2\mathcal{Q}h^s \|\mathcal{K}\|_{\infty} [2^{s\mathbf{p}} - 1]^{-\frac{1}{\mathbf{p}}}, \quad \forall n \geq 1.$$

Assumption 1 (2) implies that we can assert that $B(z, \cdot)$ is continuous on $[n^{-1}h, h]$. Hence for any $z \in \mathbb{R}$ there exists $\delta(z) \in [n^{-1}h, h]$ such that

$$(5.12) \quad \Delta_{\mathcal{K},g}^{(n)}(h, z) = B(z, \delta(z)).$$

For any $l = 0, \dots, \log_2 n - 1$ (w.l.g. $\log_2 n$ is assumed an integer) we consider the slices $V_l = \{z \in \mathbb{R} : a_{l+1} < \delta(z) \leq a_l\}$ with $a_l = 2^{-l}h$. Later on the integration over empty set is supposed to be zero. Then

$$(5.13) \quad \left\| \Delta_{\mathcal{K},g}^{(n)}(h, \cdot) \right\|_{\mathbf{p}}^{\mathbf{p}} = \sum_{l=0}^{\log_2 n - 1} \int_{V_l} |B(z, \delta(z))|^{\mathbf{p}} dz.$$

We will treat the cases $s \leq 1$ and $s > 1$ separately. If $s < 1$ on any slice V_l , $l = 0, \dots, \log_2 n$,

$$(5.14) \quad \begin{aligned} B(z, \delta(z)) &\leq \frac{\|\mathcal{K}\|_{\infty}}{\delta(z)} \int_{-\frac{\delta(z)}{2}}^{\frac{\delta(z)}{2}} |g(z+v) - g(z)| dv \leq \frac{2\|\mathcal{K}\|_{\infty}}{a_l} \int_{-\frac{a_l}{2}}^{\frac{a_l}{2}} |g(z+v) - g(z)| dv \\ &= 2\|\mathcal{K}\|_{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} |g(z+ta_l) - g(z)| dt. \end{aligned}$$

We obtain from (5.13) and (5.14) with the use of Minkowski's inequality for integrals and writing for ease of notation $\mu = 2\|\mathcal{K}\|_\infty$ that

$$\begin{aligned} \left\| \Delta_{\mathcal{K},g}^{(n)}(h, \cdot) \right\|_{\mathfrak{p}}^{\mathfrak{p}} &\leq \mu^{\mathfrak{p}} \sum_{l=0}^{\log_2 n-1} \int \left| \int_{-\frac{1}{2}}^{\frac{1}{2}} |g(ta_l + z) - g(z)| dt \right|^{\mathfrak{p}} dz \\ &\leq \mu^{\mathfrak{p}} \sum_{l=0}^{\log_2 n-1} \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \|g(\cdot + ta_l) - g(\cdot)\|_{\mathfrak{p}} dt \right)^{\mathfrak{p}} \leq \left[\frac{\mathcal{Q}h^s \|\mathcal{K}\|_\infty 2^{1-s}}{(s+1)} \right]^{\mathfrak{p}} \sum_{l=0}^{\infty} 2^{-l s \mathfrak{p}}. \end{aligned}$$

Here we have used that $g \in \mathbb{N}_{\mathfrak{p}}(s, \mathcal{Q})$. Thus, we have for any $s \leq 1$ and any $n \geq 1$

$$(5.15) \quad \sup_{g \in \mathbb{N}_{\mathfrak{p}}(s, \mathcal{Q})} \left\| \Delta_{\mathcal{K},g}^{(n)}(h, \cdot) \right\|_{\mathfrak{p}} \leq 2\mathcal{Q}h^s \|\mathcal{K}\|_\infty [2^{s\mathfrak{p}} - 1]^{-\frac{1}{\mathfrak{p}}}.$$

If $s > 1$, using Taylor's formula we have for any $g \in \mathbb{N}_{\mathfrak{p}}(s, \mathcal{Q})$ any $v \in \mathbb{R}$

$$g(v+z) - g(z) = \sum_{m=1}^{m_s} \frac{g^{(m)}(z)}{m!} v^m + \frac{v^{m_s}}{(m_s-1)!} \int_0^1 (1-\lambda)^{m_s-1} \left[g^{(m_s)}(z+v\lambda) - g^{(m_s)}(z) \right] d\lambda.$$

We have in view of Assumptions 1 and 2 for any $z \in \mathbb{R}$

$$B(z, \delta(z)) \leq \frac{\|\mathcal{K}\|_\infty}{(m_s-1)!} \frac{1}{\delta(z)} \int_{-\frac{\delta(z)}{2}}^{\frac{\delta(z)}{2}} \int_0^1 |v|^{m_s} (1-\lambda)^{m_s-1} \left| g^{(m_s)}(z+\lambda v) - g^{(m_s)}(z) \right| d\lambda dv.$$

By the latter inequality for any $z \in V_l$ we get

$$(5.16) \quad B(z, \delta(z)) \leq \frac{2\|\mathcal{K}\|_\infty a_l^{m_s}}{(m_s-1)!} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 |t|^{m_s} (1-\lambda)^{m_s-1} \left| g^{(m_s)}(z+\lambda ta_l) - g^{(m_s)}(z) \right| d\lambda dt.$$

Thus, we obtain from (5.12), (5.13) and (5.16) with the use of Minkowskii inequality for integrals and denoting $\mu = 2\|\mathcal{K}\|_\infty / (m_s-1)!$ that

$$\begin{aligned} \left\| \Delta_{\mathcal{K},f}^{(n)}(h, \cdot) \right\|_{\mathfrak{p}}^{\mathfrak{p}} &= \sum_{l=0}^{\log_2 n-1} \int_{V_l} |B(z, \delta(z))|^{\mathfrak{p}} dz \\ &\leq \mu^{\mathfrak{p}} \sum_{l=0}^{\log_2 n-1} a_l^{m_s \mathfrak{p}} \int \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 |t|^{m_s} (1-\lambda)^{m_s-1} \left| g^{(m_s)}(z+\lambda ta_l) - g^{(m_s)}(z) \right| d\lambda dt \right)^{\mathfrak{p}} dz \\ &\leq \mu^{\mathfrak{p}} \sum_{l=0}^{\log_2 n-1} a_l^{m_s \mathfrak{p}} \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 |t|^{m_s} (1-\lambda)^{m_s-1} \left\| g^{(m_s)}(\cdot + \lambda ta_l) - g^{(m_s)}(\cdot) \right\|_{\mathfrak{p}} d\lambda dt \right)^{\mathfrak{p}} \\ &\leq \left[\frac{\mathcal{Q}h^s \|\mathcal{K}\|_\infty 2^{1-s}}{(s+1)(m_s+1)(m_s-1)!} \right]^{\mathfrak{p}} \sum_{l=0}^{\infty} 2^{-l s \mathfrak{p}}. \end{aligned}$$

Here we have used that $g \in \mathbb{N}_p(s, \mathcal{Q})$. Thus, we have for any $s > 1$ and $n \geq 1$

$$(5.17) \quad \sup_{g \in \mathbb{N}_p(s, \mathcal{Q})} \left\| \Delta_{\mathcal{K}, g}^{(n)}(h, \cdot) \right\|_p \leq 2\mathcal{Q}h^s \|\mathcal{K}\|_\infty [2^{sp} - 1]^{-\frac{1}{p}}.$$

We conclude that (5.11) is established in view (5.15) and (5.17). ■

References.

- BAUER, F., HOHAGE, T. and MUNK, A. (2009). Iteratively regularized Gauss-Newton method for nonlinear inverse problems with random noise. *SIAM J. Numer. Anal.* **47:3** 1827–1846. MR2505875.
- BERTIN, K. and RIVOIRARD, V. (2009). Maxiset in sup-norm for kernel estimators. *TEST* **18:3** 475–496. MR2566412.
- BESOV, O. V., IL'IN, V. P. and NIKOL'SKII, S. M. (1978, 1979). *Integral Representations of Functions and Imbedding Theorems.*, Vol. I, II. Scripta Series in Mathematics., V. H. Winston & Sons, Washington, D.C.; Halsted Press [John Wiley & Sons], New York-Toronto, Ont.-London. MR0519341, MR0521808.
- CHICHIGNOUD, M. (2012). Minimax and minimax adaptive estimation in multiplicative regression: locally Bayesian approach. *Probab. Theory Related Fields* **153:3–4** 543–586. MR2948686.
- DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3:3** 215–228. MR1400080.
- DONOHU, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57** 301–369. MR1323344.
- GAÏFFAS, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM P & S* **11**, 334–364. MR2339297.
- GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14:4** 1150–1190. MR2543590.
- GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via L_p -norm oracle inequalities. *Probab. Theory Related Fields* **143:1–2** 41–71. MR2449122.
- GAÏFFAS S. and LECUÉ G. (2007). Optimal rates and adaptation in the single-index model using aggregation. *Electronic J. Statist.* **1**, 538–573. MR2369025.
- GOLUBEV, G. K. (1992). Asymptotically minimax estimation of a regression function in an additive model. *Problems Inform. Transmission* **28:2** 101–112. MR1178413.
- GYÖRFI, L., KOHLER, M., KRZYŻCAK, A. and WALK, H. (2002). *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York. MR1920390.
- HÄRDLE, W., MÜLLER, M., SPERLICH, S. and WERWATZ, A. (2004). *Nonparametric and semiparametric models*. Springer Series in Statistics. Springer-Verlag, New York. MR2061786.
- HOROWITZ J. L. (1998). *Semiparametric and Nonparametric Methods in Econometrics*. Lecture Notes in Statistics, 131. Springer-Verlag, New York, 1998. MR1624936.
- KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121**, 137–170. MR1863916.
- KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2008). Nonlinear estimation in anisotropic multi-index denoising. Sparse case. *Probab. Theory Appl.* **52**, 150–171. MR2354574.
- KOROSTELEV, A. and KOROSTELEVA, O. (2011). *Mathematical statistics. Asymptotic minimax theory*. Graduate Studies in Mathematics, 119. American Mathematical Society, Providence, RI. MR2767163
- LEPSKII, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35:3** 454–466. MR1091202.
- LEPSKI, O. V. and LEVIT, B. Y. (1999). Adaptive nonparametric estimation of smooth multivariate functions. *Math. Methods Statist.* **8:3** 344–370. MR1735470.

- LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25:3** 929–947. MR1447734.
- LIFSHITS M.A. (1995). *Gaussian Random Functions*. Mathematics and its Applications, 322. Kluwer Academic Publishers, Dordrecht. MR1472736.
- SERDYUKOVA, N. (2012). Spatial adaptation in heteroscedastic regression: propagation approach. *Electron. J. Stat.* **6** 861–907. MR2988432
- STONE, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13:2** 689–705. MR0790566
- WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and integral. An introduction to real analysis*. Pure and Applied Mathematics, Vol. 43. Marcel Dekker, Inc., New York-Basel. MR0492146.

LABORATOIRE D'ANALYSE,
TOPOLOGIE, PROBABILITÉS UMR 7353
AIX-MARSEILLE UNIVERSITÉ
39, RUE F. JOLIO CURIE
13453 MARSEILLE FRANCE
E-MAIL: Oleg.Lepski@cmi.univ-mrs.fr

INSTITUTE FOR MATHEMATICAL STOCHASTICS
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
GOLDSCHMIDTSTRASSE 7
37077 GÖTTINGEN GERMANY
E-MAIL: Nora.Serdyukova@gmail.com